NAVAL RESEARCH LOGISTICS QUARTERLY

OFFICE OF NAVAL RESEARCH

CONTENTS

	Page
ARTICLES	
Statistical Aspects of Collision Warning System Design Under the Assumption of Constant Velocity Courses Judah Rosenblatt	317
Lognormal Distribution and Maintainability in Support Systems Research R. L. Bovaird and H. I. Zagor	343
The Minimax Path in a Search for a Circle in a Plane Brian Gluss	357
Material Logistic Support of Weapon Systems LCDR D. W. Whelan	361
A Note on Pyramid Building Zivia S. Wurtele	377
Aircraft Compartment Design Criteria for the Army Deployment Mission W. F. Sharpe	381
A Model for Procurement, Allocation, and Redistribution for Low Demand Items G. Hadley and T. M. Whitin	395
On Minimax Surveillance Schedules C. Derman	415
NOTES	421
NEWS AND MEMORANDA	427
RECENT PUBLICATIONS	429
INDEX OF VOLUME 8	431

VOL. 8, NO. 4 DECEMBER 1961

NAVEXOS P-1278

NAVAL RESEARCH LOGISTICS QUARTERLY

EDITORS

Rear Admiral H. E. Eccles, USN (Retired)
The George Washington University

O. Morgenstern
Princeton University

F. D. Rigby
Office of Naval Research

D. M. Gilford
Office of Naval Research

Jack Laderman
Managing Editor
Office of Naval Research
Washington 25, D. C.

B. J. Beall
Secretary
Office of Naval Research
Washington 25, D. C.

ASSOCIATE EDITORS

R. Bellman, RAND Corporation
J. C. Busby, Jr., Commander, SC, USN
W. W. Cooper, Carnegie Institute of Technology
J. G. Dean, Captain, SC, USN
G. Dyer, Vice Admiral, USN (Retired)
P. L. Folsom, Captain, USN
M. A. Geisler, RAND Corporation
A. J. Hoffman, General Electric Company
H. P. Jones, Commander, SC, USN
S. Karlin, Stanford University
H. W. Kuhn, Princeton University
R. J. Lundegard, Office of Naval Research
W. H. Marlow, The George Washington University
R. E. McShane, Vice Admiral, USN (Retired)
W. F. Millson, Captain, SC, USN
H. D. Moore, Commander, SC, USN
M. I. Rosenberg, Captain, USN (Retired)
D. Rosenblatt, The George Washington University
J. V. Rosapepe, Commander, SC, USN
E. K. Scofield, Captain, SC, USN
S. M. Selig, Office of Naval Research
M. W. Shelly, Office of Naval Research
J. R. Simpson, Bureau of Supplies and Accounts
J. S. Skoczylas, Colonel, USMC
H. Solomon, The George Washington University
I. Stakgold, Northwestern University
E. D. Stanley, Rear Admiral, SC, USN
C. Stein, Jr., Captain, SC, USN (Retired)
R. M. Thrall, University of Michigan
C. B. Tompkins, University of California
J. F. Tynan, Commander, SC, USN
J. D. Wilkes, Office of Naval Research

The Naval Research Logistics Quarterly is devoted to the dissemination of scientific information in lo tics and will publish research and expository papers, including those in certain areas of mathemat statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness logistics operations.

Information for Contributors is indicated on inside back cover.

The Naval Research Logistics Quarterly is published by the Office of Naval Research in the month March, June, September, and December and can be purchased from the Superintendent of Documents, Government Printing Office, Washington 25, D.C. Subscription Price: $1.50 a year in the U.S. and Can $2.00 elsewhere; $0.50 for a single copy. Reprints of an individual article can be purchased in quant of 100 or more. Requests for the purchase price of reprints of a particular article should be sent to Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C.

The views and opinions expressed in this quarterly are those of the authors and not necessarily tho the Office of Naval Research.

The issuance of this publication approved by the Secretary of the Navy on 24 March 1961.

Permission has been granted to use the copyrighted material appearing in this publication.

STATISTICAL ASPECTS OF COLLISION WARNING SYSTEM DESIGN* †
UNDER THE ASSUMPTION OF CONSTANT VELOCITY COURSES

Judah Rosenblatt

University of New Mexico

INTRODUCTION

The increase in aircraft speed, due to the advent of high-speed jet airplanes, has made it necessary to devise superior ways of avoiding mid-air collisions. Because so much distance is covered in the time it takes an aircraft to react to its controls, any maneuvers taken to avoid a collision must be executed well in advance of such a possibility. For this reason, a visual determination of the likelihood of collision is no longer practical. In a previous article, Pina and Landau [1], have proposed using radar equipment, and devised a scheme for taking and processing the observations to achieve a useful collision warning system.

The purpose of this article is to find the best procedure for taking and processing these observations in a manner consistent with certain restraints, namely a limit on the number of available observations, which results from the limited detection range, and the existence of a required warning time.

This article attempts to refine both the method of taking observations and the method of processing observations which have previously been proposed.

Under the idealized assumptions made, these results are an extension and refinement of the work of Pina and Landau [1]. Even if the system proposed here is not used, possibly due to obvious weight restrictions, it will serve as a yardstick by which the performance of any proposed system can be measured.

ASSUMPTIONS

1. The two airplanes of interest each fly constant velocity (straight line, constant speed). (Thus, the important case of flight in the vicinity of aircraft terminals is not handled in this article.)
2. The first airplane, P_1, is equipped with the complete collision warning system, and the second airplane, P_2, has at least a responder.
3. P_1 can take measurements on range. Measurements on range are considered to have maximum error of σ_0 feet. (In the Boeing system $\sigma_0 = 50$, for reasons to be given.)

*Manuscript received May 8, 1961.

† This study was initiated during the summer of 1957 while the author was a summer faculty associate in what is presently the Analysis Unit, Engineering Computing and Analysis Staff, Transport Division, Boeing Airplane Company, and supported by Boeing until the major results were obtained.

4. P_1 can take measurements on bearing angle. Bearing angle measurements are assumed to be normally distributed, with mean value equal to the true bearing angle, and standard deviation of σ radians. (In the Boeing system $\sigma = 0.00087$. This is an attempt to realize mathematically the condition that maximum error in bearing angle measurement is supposed to be 0.15 degrees = 0.00261 radians.)
5. All measurement errors are (mutually) statistically independent.
6. If the antenna of P_1 faces P_2, and if the range does not exceed some given range S, the probability of P_1 detecting P_2 is unity. (The results of this assumption can easily be modified to handle the case where "Probability of Detection" versus "Range" is known.)
7. There is one complete antenna revolution every t_o seconds (in the Boeing system $t_o = 4$), with one measurement taken on each antenna revolution.
8. In the event of a warning, a maneuver to shift the miss distance by $\mathfrak{Z}$ feet is to be executed (in the Boeing system $\mathfrak{Z} = 1000$), at least t_w seconds before the impending collision, t_w chosen so this maneuver can be executed, (in the Boeing system $t_w = 12.5$).

Some discussion of these assumptions is in order. While it is true that even if it is attempted, airplanes cannot fly a perfect constant velocity course, it is felt that the deviations from constant velocity can be held well below the error in measurement.

The reasons for not using a simple reflection radar (i.e., one based on the reflection of transmitted waves, much like a flashlight) are:

Since the transmitted wave need only be strong enough to be detected by the oncoming airplane (rather than so strong that its reflection from the oncoming airplane be detectable at the transmitter), the range in which detection is likely is increased. Measurement error is greatly reduced, since now all measurements are from antenna to antenna.

The maximum error in distance measurement is derived from the error in determining the time when the signal from the responder is first received. It takes a known finite time for a pulse to rise to its maximum value. It is assumed that the pulse will surely be detected by the time it has reached this maximum value, and will not be detected before it starts its rise. For this reason, the error in measuring time of arrival cannot exceed this known "rise time." From the maximum error in measurement of time of arrival, and knowledge of the speed of radio waves, it is simple to determine the maximum error in distance measurement.

The normality assumption, concerning the bearing-angle measurements has some basis in theory (Central Limit Theorem), but is open to serious question. However, results based on "normal theory" often hold up well when this assumption is violated. Further, this theory is mathematically tractable. In any case, before the proposed scheme should be adopted, it would be worthwhile to investigate its performance under deviations from the "normality" assumption, possibly using a Monte Carlo approach.

DISCUSSION OF RESULTS

It is desired that an alarm be sounded in P_1 if P_2 will approach within some predetermined distance, $|d_0|$, of P_1, and that an alarm be sounded at least t_w seconds before such an occurrence. If P_2 will never approach within $|d_0|$ of P_1 it is desired that no alarm be sounded. The distance $|d_0|$ is chosen so that if the minimum miss distance exceeds it, a collision is impossible, taking into account the finite size of the airplanes. The technique suggested here allows (to within high accuracy) the probability of sounding an alarm when minimum miss distance is $|d_0|$ to be set at any desired level $1 - \alpha$.

Because of errors in measurement, it is often impossible to achieve the desired results. The achievable results are computable as a one parameter family of graphs (functions) of "Probability of Sounding an Alarm" versus "Minimum Miss Distance." The parameter is $|v|$, the magnitude of the relative velocity.

If the assumption of normality of bearing-angle measurements is reasonable, the theory developed here is a refinement of that of Pina and Landau [1]. It yields improvement where most needed, at the expense of requiring a more complicated computer and greater storage capability. Compared to the Pina-Landau results, this procedure improves the ability to discriminate between collision and non collision courses. (Note that maneuvers, in the case of a course far from collision, can be quite dangerous.) Such improvement in the performance curve is to be expected, since this procedure assumes normality, while the Pina-Landau procedure leaves the distribution unspecified, and puts only a restriction on its range.

NOTATION

Capital letters with a $\wedge$ on top of them denote random variables.

The corresponding capital letters (without a $\wedge$) denote the values taken on by the random variables.

The corresponding small letters denote "true" values.

Thus, if $\hat{R}$ is a random variable corresponding to range measurement, and $\hat{X}$ is one corresponding to bearing angle measurement, particular observed values of $\hat{R}$ and $\hat{X}$ are denoted by R, X. The true range and bearing angle are denoted by r, x. From numbers 3 and 4 in the section on Assumptions, we conclude that

$$P\left\{|\hat{R} - r| \leq \sigma_0\right\} = 1 \tag{1}$$

and

$$E(\hat{X}) = x. \tag{2}$$

As noted in number 4 in the section on Assumptions, $\hat{X}$ is assumed to be normally distributed with standard deviation σ.

THEORY NEGLECTING ERRORS IN RANGE MEASUREMENT

All of this analysis will be done relative to P_1, that is with P_1 considered as a fixed point, and it will be a two-dimensional analysis. This is reasonable, since though it is azimuth and elevation that may be measured, these can be converted to a "measurement" of bearing angle. For details, see Pina-Landau [1], pp. 796-798.

We shall postpone until a later section (The Sampling Plan — Which Measurement to Store and When to Cease Taking Samples) the questions of when to cease taking observations and which observations are to be stored for use in making the final decision, and make

Assumption I

We shall assume that we are given the measurements $R_1, \ldots, R_n$ on range, and the corresponding measurements $X_1, \ldots, X_n$ on bearing angle, which are to be used in making the final decision, and that these measurements are ordered in time. We shall further assume that for given ranges $r_1, \ldots, r_n$ and first bearing angle x_1, (i.e., given the parameters which specify completely the true geometric situation with respect to the reference frame), $(\hat{R}_1, \ldots, \hat{R}_n)$ and $(\hat{X}_1, \ldots, \hat{X}_n)$ are statistically independent of each other, and $(\hat{X}_1, \ldots, \hat{X}_n)$ are mutually statistically independent.

We shall indicate in the section dealing with the sampling plan how to satisfy Assumption I. Assumption I simply means that if $r_1, \ldots, r_n$, x_1 are given, the knowledge of $R_1, \ldots, R_n$ does not change the statistical description of $\hat{X}_1, \ldots, \hat{X}_n$, and knowledge of any of the X_i does not alter the statistical description of any of the other $\hat{X}_j$. Clearly, we must be given x_1 for the above assumption to be reasonable, due to the arbitrariness of the reference frame, for without x_1 given, clearly the $\hat{X}_i$ could not be independent. As is seen, the performance of the test to be developed in no way depends on x_1.

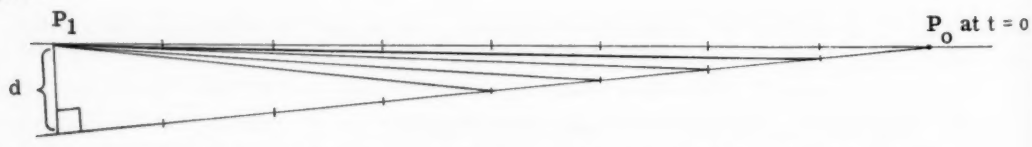
It is intuitively seen that the only relevant information provided by the X_i is given by

$$\hat{Y}_i = \hat{X}_{i+1} - \hat{X}_i, \quad i = 1, \ldots, n-1 \tag{3}$$

(Y_i being the measured bearing angle changes), since bearing angle is measured relative to an arbitrary reference frame. This intuitive reasoning may be further justified to believers in Likelihood Ratio Tests, by showing the test about to be derived, to be the Likelihood Ratio Test of the hypothesis $d = 0$ against the hypothesis $d \neq 0$, based on $X_1, \ldots, X_n$, where

$$|d| = \text{minimum } |\text{miss distance}|. \tag{4}$$

Assuming $r_i > r_{i+1}$, an assumption justified in the section dealing with the Sampling Plan, it can easily be seen from paragraph 1 of the section on Assumptions, that if three different ranges, together with their corresponding times, are known perfectly, then $|d|$ can be computed perfectly. Thus, we might consider a collision warning system based on range and time measurements only. However, as shown in the following illustration, a change in $|d|$ from 0 to 1000 feet, only alters those ranges in excess of 15,000 feet by at most 9 feet, for given



first range of 60,000 feet. Thus, an error in range measurement of 9 feet could alter the computed value of $|d|$ by 1000 feet, making a collision warning system based on range measurements alone quite useless under number 3 of the section on Assumptions.

Thus, it appears that measured range values alone furnish us with practically none of the desired information. However, from

$$y_i \equiv x_i - x_{i+1} = \arcsin \frac{d}{r_{i+1}} - \arcsin \frac{d}{r_i} \doteq d\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right), \tag{5}$$

we see that for $r_i - r_{i+1}$ sufficiently large, a knowledge of the bearing angle change together with knowledge of the corresponding ranges should enable accurate computation of d. Thus for the purposes of this section, we shall make Assumption II.

Assumption II

$r_1, \ldots, r_n$ are known perfectly. A collision warning system based on $X_1, \ldots, X_n$ using $r_1, \ldots, r_n$ only to find d for the measured $X_1, \ldots, X_n$, using the approximation in Eq. (5) will be derived. We shall later determine the effect of using the approximation (5). In the next section we shall relax Assumption II, i.e., in the test procedure replace $r_1, \ldots, r_n$ by the measured ranges $R_1, \ldots, R_n$, and we shall show that such a modification has very little effect on the results derived in this section.

Under Assumption I, for given $|v|$, $r_1, \ldots, r_n$, x_1, the $\hat{X}_i$, $i = 1, \ldots, n$ are statistically independent normal (Gaussian) random variables with mean values x_i and common variance σ^2. Thus, the random vector

$$\hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_{n-1} \end{pmatrix} \tag{6}$$

has a multivariate normal distribution with mean vector, assuming the approximation (5),

$$E_d(\hat{Y}) = d \begin{pmatrix} \frac{1}{r_2} - \frac{1}{r_1} \\ \vdots \\ \frac{1}{r_n} - \frac{1}{r_{n-1}} \end{pmatrix} \equiv d\,c \tag{7}$$

and covariance matrix

$$A^{(n-1)\times(n-1)} = \begin{bmatrix} 2\sigma^2 & -\sigma^2 & 0 & . & . & . & 0 \\ -\sigma^2 & 2\sigma^2 & -\sigma^2 & . & . & . & 0 \\ 0 & -\sigma^2 & 2\sigma^2 & -\sigma^2 & . & . & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & . & . & . & 0 & -\sigma^2 & 2\sigma^2 \end{bmatrix}, \tag{8}$$

i.e.,

$$(A)_{ij} = \begin{cases} 2\sigma^2 & \text{if } i = j, \\ -\sigma^2 & \text{if } j = i+1,\ j = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the density $f_{\hat{Y}}(\,\cdot\,; d)$ of $\hat{Y}$ is given by

$$f_{\hat{Y}}(\lambda; d) = \mathcal{K} e^{-\frac{1}{2}(\lambda - dc)' A^{-1} (\lambda - dc)}, \tag{9}$$

where

$$\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_{n-1} \end{pmatrix}$$

and the constant κ can be chosen independent of d so that

$$\int_{-\infty}^{\infty} \cdots \int f_{\hat{Y}}(\lambda; d)\, d\lambda = 1.$$

With a slight bit of manipulation, we see that

$$f_{\hat{Y}}(\lambda; d) = \kappa e^{-\frac{1}{2}\lambda' A^{-1}\lambda} \cdot e^{\frac{1}{2}\left[2d\lambda' A^{-1}c - d^2 c' A^{-1}c\right]} \equiv g(\lambda) \cdot h[z(\lambda); d], \tag{10}$$

where

$$z(\lambda) = \lambda' A^{-1} c\,. \tag{11}$$

Making use of the Neyman factorization theorem, we see that the random variable $z(\hat{Y})$ is sufficient for d (Lehmann [2].) What this means is that all of the information present in $\hat{Y}$ concerning d is contained in the random variable $z(\hat{Y})$.

Since

$$(A^{-1})_{ij} = \begin{cases} i(n-j) & \text{if } i \le j, \\ j(n-i) & \text{if } i > j, \end{cases} \qquad i, j = 1, \ldots, n-1, \tag{12}$$

we see that

$$z(\hat{Y}) = \sum_{i=1}^{n-1} \left[(n-i) \sum_{j=1}^{i} j\left(\frac{1}{r_{j+1}} - \frac{1}{r_j}\right) + i \sum_{j=i+1}^{n} (n-j)\left(\frac{1}{r_{j+1}} - \frac{1}{r_j}\right) \right] \hat{Y}_i$$

$$\equiv \sum_{i=1}^{n-1} a_i^* \hat{Y}_i\,. \tag{13}$$

Note that $a_i^* > 0$ if $r_{j+1} < r_j$ (see the section on the sampling plan, where this is justified). Since any continuous one-one function of a sufficient statistic for d is sufficient, it is preferable to work with

$$(14)\qquad \sum_{i=1}^{n-1} a_i \hat{Y}_i \equiv \sum_{i=1}^{n-1} \frac{a_i^*}{\sqrt{2}\,\sigma\sqrt{\sum_{j=1}^{n-1} a_j^{*2} - \sum_{j=1}^{n-2} a_j^* a_{j+1}^*}}\,\hat{Y}_i,$$

which is a normal random variable with mean value proportional to d, (assuming the approximation in Eq. 5), and unit variance. This follows, since for any c_i

$$\operatorname{Var} c_i \hat{Y}_i = 2\sigma^2 \left(\sum_{i=1}^{n-1} c_i^2 - \sum_{i=1}^{n-2} c_i c_{i+1} \right) > 0$$

if not all c_i are 0, since

$$\sum_{i=1}^{n-1} c_i^2 - \sum_{i=1}^{n-2} c_i c_{i+1} = \frac{1}{2} \sum_{i=0}^{n-1} (c_{i+1} - c_i)^2$$

with $c_0 = c_n = 0$. Thus, $a_i > 0$. A reasonable test of the hypothesis $\mathcal{H}_0$ that $|d| \leq |d_0|$ against the hypothesis $\mathcal{H}_1$ that $|d| > |d_0|$, of level α is given by the procedure: assert that $\mathcal{H}_0$ is true if and only if

$$(15)\qquad \left| \sum_{i=1}^{n-1} a_i Y_i \right| \leq c_\alpha,$$

where c_α is chosen so that

$$P_{|d| = |d_0|} \left\{ \left| \sum_{i=1}^{n-1} a_i \hat{Y}_i \right| \leq c_\alpha \right\} = 1 - \alpha.$$

(Note that c_α is dependent on $|d_0|$, though this is not explicit in the notation.) This test procedure is optimal, under the approximations in Eq. (5), in the sense of being a uniformly most powerful unbiased test of $\mathcal{H}_0$ against $\mathcal{H}_1$ of level α (Lehmann) [2], but this optimality will be of little concern to us here, and we will try to justify the whole procedure (including our approximations) by comparing this test's performance with that of certain tests which are optimal in a much more desirable sense to be mentioned soon.

To accomplish what is desired, let us examine the simpler situation in which we assume that there are only two possible miss distances $|d_0|$ and $|d_1| > |d_0|$, with $\hat{Y}$ having density f_s when $d = |d_s|$, $(s = 0, 1)$.

To test the hypothesis $\mathcal{H}_0'$: $d = |d_0|$ against the hypothesis $\mathcal{H}_1'$: $d = |d_1|$, we make use of a theorem of Neyman and Pearson (Lehmann, [2]) which states that under certain mild mathematical restrictions (which are here satisfied), there is a class of tests of a simple hypothesis $\mathcal{H}_0^*$ (i.e., a hypothesis which uniquely specifies a density) against the simple hypothesis $\mathcal{H}_1^*$, for which the one satisfying

$$(1') \qquad P_{\mathcal{H}_0^*}\{\text{Reject } \mathcal{H}_0^*\} = \alpha$$

is the unique test minimizing

$$(2') \qquad P_{\mathcal{H}_1^*}\{\text{Reject } \mathcal{H}_1^*\},$$

of all tests satisfying (1').

No test of size α can hope to do better in deciding between $\mathcal{H}_0^*$ and $\mathcal{H}_1^*$. We note that if for each $\alpha \in (0,1)$ there is a Neyman-Pearson test, then since any test T of $\mathcal{H}_0^*$ versus $\mathcal{H}_1^*$ must be of some size, there is always a Neyman-Pearson test which does better than T unless T is a Neyman-Pearson test.

The Neyman-Pearson Lemma is applicable to the problem of testing $\mathcal{H}_0'$ against $\mathcal{H}_1'$ because each of these hypotheses specifies one unique density for $\hat{Y}$.

The Neyman-Pearson tests of $\mathcal{H}_0'$ against $\mathcal{H}_1'$ are those for which we accept $\mathcal{H}_0'$ if and only if

$$\frac{f_1(Y_1,\ldots,Y_n)}{f_0(Y_1,\ldots,Y_n)} \leq k,$$

where k is some nonnegative real number. Under the assumption of normality of bearing measurements the density of $\hat{Y}$ under $\mathcal{H}_i'$ is normal, with mean vector

$$E_{\mathcal{H}_i'}(\hat{Y}) \equiv \begin{pmatrix} y_{i,1} \\ \vdots \\ y_{i,n-1} \end{pmatrix}$$

and covariance matrix A as given in Eq. (8). Omitting the rather detailed computations involved, we can show that in this case the Neyman-Pearson tests of $\mathcal{H}_0'$ versus $\mathcal{H}_1'$ are those for which we accept $\mathcal{H}_0'$ if and only if

$$(16) \quad \sum_{i=1}^{n-1}\left[(n-i)\sum_{j=1}^{i} j(y_{1,j}-y_{0,j}) + i\sum_{j=i+1}^{n-1}(n-j)(y_{1,j}-y_{0,j})\right]Y_i \equiv \sum_{i=1}^{n-1} b_i^* Y_i \leq k',$$

where k' is some real number.

Again, rather than work with

$$\sum_{i=1}^{n-1} b_i^* \hat{Y}_i,$$

we choose to use the random variable,

$$\sum_{i=1}^{n-1} b_i \hat{Y}_i = \sum_{i=1}^{n-1} \frac{b_i^*}{\sqrt{2}\,\sigma \sqrt{\sum_{j=1}^{n-1} b_j^{*2} - \sum_{j=1}^{n-2} b_j^* b_{j+1}^*}} \hat{Y}_i,$$

which is a normal random variable with unit variance.

We now show that under the exact part of Eq. (5) the test given in Eq. (15) does almost exactly as well as the Neyman-Pearson test in distinguishing between certain crucial values of $|d_0|$ and $|d_1|$ in those cases where the Neyman-Pearson test does well in such distinguishing. Letting

$$a_i'^* = \sum_{s=0}^{1} (-1)^{s-1} \left[(n-i) \sum_{j=1}^{i} j \left(\arcsin \frac{|d_s|}{r_{j+1}} - \arcsin \frac{|d_s|}{r_j} \right) \right.$$

$$\left. + i \sum_{j=i+1}^{n-1} (n-j) \left(\arcsin \frac{|d_s|}{r_{j+1}} - \arcsin \frac{|d_s|}{r_j} \right) \right], \tag{17}$$

and defining a_i' in the same manner as b_i was defined from b_i^* after Eq. (16), then from Eq. (16) the Neyman-Pearson test of size α of $\mathcal{H}_0'$ versus $\mathcal{H}_1'$, assuming the exact formulas of Eq. (5) is to accept $\mathcal{H}_0'$ if and only if

$$\sum_{i=1}^{n-1} a_i' Y_i \leq c_\alpha', \tag{18}$$

where c_α' is chosen so that

$$P_{\mathcal{H}_0'} \left\{ \sum_{i=1}^{n-1} a_i' \hat{Y}_i \leq c_\alpha' \right\} = 1 - \alpha$$

(this latter being computed under the exact part of Eq. (5). Note that c_α' depends on d_0 and d_1, since a_i' depends on them. The desired result will be demonstrated in two rather complicated steps, all computations being done under the exact formulas of Eq. (5). In the first step, we show for the crucial values of $|d_0|$ and $|d_1|$ that the test given in Eq. (15) does almost as well in distinguishing between $|d_0|$ and $|d_1|$ as the test:

$$\text{accept } \mathcal{H}_o' \text{ if and only if } \sum_{i=1}^{n-1} a_i Y_i \leq \tilde{c}_\alpha,$$

(19) $$\text{where } P_{\mathcal{H}_o'} \left\{ \sum_{i=1}^{n-1} a_i \hat{Y}_i \leq \tilde{c}_\alpha \right\} = 1 - \alpha.$$

In the second step, we show for the crucial values of $|d_o|$ and $|d_1|$ that this latter test does almost as well in distinguishing between $|d_o|$ and $|d_1|$ as the Neyman-Pearson test Eq. (18).

In order to find the crucial values of $|d_o|$ and $|d_1|$, we assume that the sum of the wingspans of P_1 and P_2 is 2ω feet. Since the maneuver in case of impending collision involves a shift in miss distance of 1000 feet, it is vital that a useful test distinguish well between

$$|d_o| = \omega \quad \text{and} \quad |d_1| = 1000 - \omega.$$

For modern high-speed aircraft we may assume $100 \leq \omega \leq 300$. Thus, the crucial values satisfy

(20) $$|d_1| < 9|d_o|$$

and

(21) $$\frac{7}{3}|d_o| < |d_1|.$$

Since it is clear that a miss distance greater than $d > 0$ has a higher probability of being interpreted as a noncollision than d, we will restrict our attention to the cases

(22) $$0 \leq d \leq 1000.$$

(Completely similar results hold if we examine $d = -|d_o|$ versus $d = -|d_1| < -|d_o|$, and restrict to $-1000 \leq d \leq 0$.)

Also, since we are primarily interested in high relative speed cases, due to the required 12.5-second warning time, we restrict consideration to the cases

(23) $$r_i \geq 11{,}000.$$

We note that for $0 < x < 1$

(24) $$\arcsin x = x + \frac{1}{6} \sum_{k=0}^{\infty} c_k x^{3+2k}, \quad 0 \leq c_k \leq 1.$$

(See Courant, [3], Vol. 1, p. 408.) Thus, if $1 > x > y > 0$

$$(25) \qquad x-y \le \arcsin x - \arcsin y \le x-y + \frac{1}{6} \sum_{k=0}^{\infty} \left(x^{3+2k} - y^{3+2k}\right)$$

$$= (x-y)\left[1 + \frac{1}{6} \frac{x^2 + xy + y^2 - x^2y^2}{(1-x^2)(1-y^2)}\right].$$

Now making use of Eqs. (22), (23), and (25), we find

$$(26) \qquad \left(\frac{d}{r_{i+1}} - \frac{d}{r_i}\right) \le \arcsin \frac{d}{r_{i+1}} - \arcsin \frac{d}{r_i} \le \left(\frac{d}{r_{i+1}} - \frac{d}{r_i}\right) 1.0042 .$$

Now, letting

$$\gamma = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_{n-1} \end{pmatrix},$$

we let

$$(27) \qquad g(\gamma\,; d) = E_d \left[\sum_{i=1}^{n-1} \frac{\gamma_i}{\sqrt{2}\sigma \sqrt{\sum_{j=1}^{n-1} \gamma_j^2 - \sum_{j=1}^{n-2} \gamma_j \gamma_{j+1}}} \hat{Y}_i \right]$$

with the expectation taken under the assumption of the exact formula for bearing angle change, i.e.,

$$E_d(\hat{Y}_i) = \arcsin \frac{d}{r_{i+1}} - \arcsin \frac{d}{r_i},$$

and let $h(\gamma\,; d)$ be the same expectation under the assumption,

$$(28) \qquad E_d(\hat{Y}_i) = d\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right).$$

In order for the Neyman-Pearson test Eq. (18) to do well for the crucial values of $|d_o|$ and $|d_1|$, since

$$\sum_{i=1}^{n-1} a_i' \hat{Y}_i$$

is a normal random variable with unit variance, it is necessary that

$$(29) \qquad g\left(a'; |d_1|\right) - g\left(a'; |d_o|\right) \ge 5$$

(where a' has the obvious meaning), as a glance at the standard normal distribution will show.

Rather indirectly (due to the extreme difficulty of working directly with the a_i' and a_i), we shall prove a similar inequality for $g(a; |d_1|) - g(a; |d_0|)$.

Noting that the values of g and h are independent of magnification of γ, we see from the stated optimality of the Neyman-Pearson test, and the fact that both

$$\sum_{i=1}^{n-1} a_i \overset{\Diamond}{Y}_i \text{ and } \sum_{i=1}^{n-1} a_i' \hat{Y}_i$$

are normal random variables with unit variance, that

(30) $$\gamma = a' \equiv \begin{pmatrix} a_1' \\ \vdots \\ a_{n-1}' \end{pmatrix} \text{ maximizes } g(\gamma; |d_1|) - g(\gamma; |d_0|).$$

while

$$\gamma = a \equiv \begin{pmatrix} a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} \text{ maximizes } h(\gamma; |d_1|) - h(\gamma; |d_0|).$$

Thus,

$$g(a; |d_1|) - g(a; |d_0|) \leq g(a'; |d_1|) - g(a'; |d_0|)$$

(31) $$\leq 1.0042\, h(a'; |d_1|) - h(a'; |d_0|) \leq 1.0074\, [h(a'; |d_1|) - h(a'; |d_0|)]$$

$$\leq 1.0074\, [h(a; |d_1|) - h(a; |d_0|)] \leq 1.0074\, [g(a; |d_1|) - g(a; |d_0|)],$$

i.e.,

$$g(a; |d_1|) - g(a; |d_0|) \leq g(a'; |d_1|) - g(a'; |d_0|) \leq 1.0074\, [g(a; |d_1|) - g(a; |d_0|)].$$

Here, the first inequality follows from (30), the second inequality from (26), the third inequality from (28) (linearity of h in d) and (21), the fourth inequality from (30), and the fifth inequality from (27), (28), and (24), because the derivative w.r.t. d of $h(a; d) - h(a; |d_0|)$ does not exceed that of $g(a; d) - g(a; |d_0|)$ for $d \geq |d_0|$ while these two expressions are equal for $d = |d_0|$.

Hence, we have from (29) and (31):

(32) $$g(a; |d_1|) - g(a; |d_0|) \geq 5/1.0074.$$

Now,

(33) $$g(a; |d_1|) \leq 1.0042\, h(a; |d_1|) \leq 9 \cdot 1.0042\, h(a; |d_0|) \leq 9.038\, g(a; |d_0|),$$

the first inequality following from (26), the second from (20) and the third from (26), all using (27) and (28). Now (33) and (32) imply

(34) $$g(a; |d_0|) \geq 5/8.1.$$

But Eq. (34) proves that the tests given by (15) and (19) are just about identical in distinguishing the crucial $|d_0|$ and $|d_1|$ values, as yet another glance at the standard normal distribution shows.

From (31), we can also carry through the final portion of the "near optimality" proof, since the significance of (31) is that for the important $|d_s|$ the statistic

$$\sum_{i=1}^{n-1} a_i \hat{Y}_i .$$

can discriminate near $|d_1|$ almost as well as can the optimum statistic

$$\sum_{i=1}^{n-1} a_i' \hat{Y}_i .$$

To be precise, the size α test of $\mathcal{H}_0'$, (19), based on

$$\sum_{i=1}^{n-1} a_i \hat{Y}_i \text{ at } d = |d_0| + 1.0074\,[\,|d_1| - |d_0|\,],$$

has at least as great a probability of choosing $\mathcal{H}_1'$ as does the Neyman-Pearson test, (18), based on

$$\sum_{i=1}^{n-1} a_i' \hat{Y}_i \text{ at } d = |d_1| ; \text{ i.e.,}$$

at worst there is a "loss of vision" of $0.0074\ [\,|d_1| - |d_0|\,]$ feet. This follows because from (31)

$$g(a'; |d_1|) - g(a'; |d_0|) \leq 1.0074[h(a; |d_1|) - h(a; |d_0|)].$$

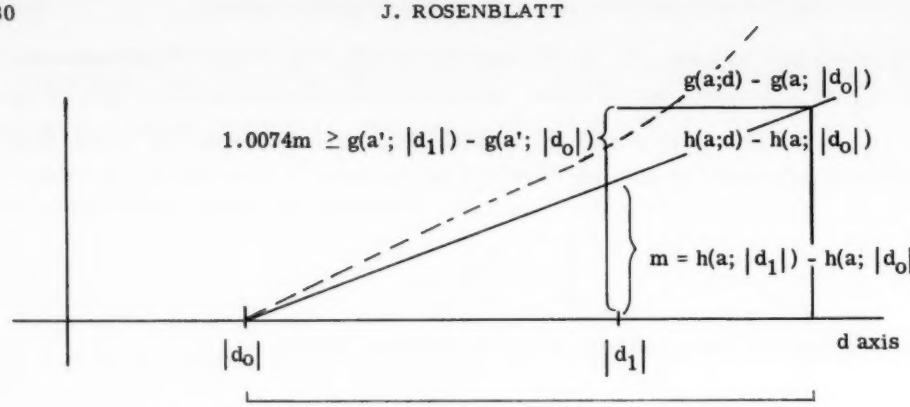
But, since $h(a;d) - h(a; |d_0|)$ is linear in $d - |d_0|$, at $d = |d_0| + 1.0074[\,|d_1| - |d_0|\,]$,

$$h(a;d) - h(a; |d_0|) \geq g(a'; |d_1|) - g(a'; |d_0|).$$

Hence, since from (24) the derivative of $g(a;d) - g(a; |d_0|)$ w.r.t. d is at least as great as that of $h(a;d) - h(a; |d_0|)$, these two expressions being equal for $d = |d_0|$, we have at $d = d_0 + 1.0074\,[\,|d_1| - |d_0|\,]$

$$g(a;d) - g(a; |d_0|) \geq h(a;d) - h(a; |d_0|)$$

from which the desired result follows. It is shown in the following illustration



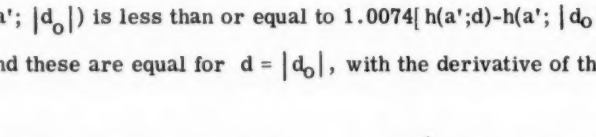
We can now prove that for all d satisfying $|d_o| \le d \le |d_1|$, with $|d_s|$ satisfying (21), g(a;d) - g(a; $|d_o|$) is close to g(a';d) - g(a'; $|d_o|$). (Note that a and a' are functions of $|d_o|$, $|d_1|$ satisfying (21)). This computation is done in three stages:

(a) We show g(a';d) - g(a'; $|d_o|$) close to h(a';d) - h(a'; $|d_o|$).

From (24) we see that the derivative of h(a';d) - h(a'; $|d_o|$) w.r.t. d does not exceed that of g(a';d) - g(a'; $|d_o|$). Since these two expressions equal 0 at $d = |d_o|$, we conclude that

(35) $$h(a';d) - h(a'; |d_o|) \le g(a';d) - g(a'; |d_o|) \text{ for } d \ge |d_o|.$$

From (24) we see also that the derivative of g(a';d) - g(a'; $|d_o|$) is increasing. Thus, since g(a';d) - g(a'; $|d_o|$) is less than or equal to 1.0074[h(a';d)-h(a'; $|d_o|$)] for $d = |d_1|$ (by (31)) and these are equal for $d = |d_o|$, with the derivative of the latter being constant



it follows that

(36) $$g(a';d) - g(a'; |d_o|) \le 1.0074\,[\,h(a';d) - h(a'; |d_o|)], \text{ for } d \ge |d_o|.$$

(b) We show h(a';d) - h(a'; $|d_o|$) close to h(a;d) - h(a; $|d_o|$).

From (31), g(a; $|d_1|$) - g(a; $|d_o|$) $\le 1.0074\,[\,h(a'; |d_1|) - h(a'; |d_o|)]$

$\le 1.0074\,[\,h(a; |d_1|) - h(a; |d_o|)] \le 1.0074\,[\,g(a; |d_1|) - g(a; |d_o|)]$.

But from (28), we see that we can write

$$h(\gamma; |d_1|) - h(\gamma; |d_o|) \equiv [\,|d_1| - |d_o|\,]\ \tau(\gamma).$$

Hence,

$$h(a';d) - h(a'; |d_o|) \le h(a;d) - h(a; |d_o|) \le 1.0074\,[h(a';d) - h(a'; |d_o|)], \quad \text{for } d \ge |d_o|. \tag{37}$$

(c) We show $h(a;d) - h(a; |d_o|)$ close to $g(a;d) - g(a; |d_o|)$.

By the same reasoning as in (a),

$$h(a;d) - h(a; |d_o|) \le g(a;d) - g(a; |d_o|) \le 1.0074\,[h(a;d) - h(a; |d_o|)]. \tag{38}$$

Combining (35), (36), (37), (38), we conclude that

$$g(a;d) - g(a; |d_o|) \le (1.0074)^2[g(a';d) - g(a'; |d_o|)] \le (1.0074)^3[g(a;d) - g(a; |d_o|)]. \tag{39}$$

In a fashion similar to (31), (39) asserts that the properties of the tests based on

$$\sum_{i=1}^{n-1} a_i \hat{Y}_i \quad \text{and} \quad \sum_{i=1}^{n-1} a_i' \hat{Y}_i$$

are similar for all $d \ge |d_o|$.

We summarize the results of this section. Based on certain approximations, we derived a test of H_o: $|d| \le |d_o|$ versus H_1: $|d| > |d_o|$ which was shown to be almost as good in distinguishing H_o': $d = |d_o|$ versus H_1': $d = |d_1|$ as the best that can be done using no approximations. Thus, the characteristics of this test can be used as a yardstick against which the performance of any test can be compared, when the model is valid.

In practice, of course, this test cannot be performed, since it assumed perfect knowledge of the ranges r_i. In the next section, we shall show that the performance is not greatly altered when the r_i are replaced by the available measurements R_i. That is, we present a performable test whose properties are close to the unachievable optimum.

REPLACING THE SUGGESTED TEST BY A PERFORMABLE TEST

The test procedure given by (15) is not a workable one, because the r_i are not known perfectly. As was pointed out, a test based on the r_i alone would be extremely sensitive to errors in the knowledge of them. In this section, we shall investigate the effect of replacing the r_i by the measured R_i in the test given by (15). Let

$$\hat{A}_i^* = (n-i) \sum_{j=1}^{i} j \left(\frac{1}{\hat{R}_{j+1}} - \frac{1}{\hat{R}_j} \right) + i \sum_{j=i+1}^{n-1} (n-j) \left(\frac{1}{\hat{R}_{j+1}} - \frac{1}{\hat{R}_j} \right) \tag{40}$$

and

$$\hat{A}_i = \frac{\hat{A}_i^*}{\sqrt{2}\,\sigma\sqrt{\sum_{j=1}^{n-1} \hat{A}_i^{*2} - \sum_{j=1}^{n-2} \hat{A}_i^* \hat{A}_{i+1}^*}} \quad . \tag{41}$$

The test procedure actually proposed to handle the collision warning problem is to accept $\mathcal{H}_o$ if and only if

$$\left| \sum_{i=1}^{n-1} A_i Y_i \right| \leq C_\alpha \,, \tag{42}$$

where A_i is the value assumed by $\hat{A}_i$ when $\hat{R}_j$ assumes the value R_j, and where, to be explicit, for the given R_i, C_α is chosen from the relation

$$P\left\{ \left| \hat{\delta} + \sum_{i=1}^{n-1} A_i Y_{oi} \right| \leq C_\alpha \right\} = 1 - \alpha \,, \tag{43}$$

with $\hat{\delta}$ a standard normal random variable and

$$Y_{oi} = |d_o| \left(\frac{1}{R_{i+1}} - \frac{1}{R_i} \right) .$$

(Note that in (43) we cannot replace Y_{oi} by y_{oi}, as would be desired, since y_{oi}, the true bearing-angle changes when $d = |d_o|$, are unknown. Note also that C_α is the value of a random variable because it depends on the values A_i and Y_{oi} of the random variables $\hat{A}_i$ and $\hat{Y}_{oi}$.) As an alternative to (42) if we considered rejecting $\mathcal{H}_o$ when $d = |d_o|$, s times as serious as accepting $\mathcal{H}_o$ when $d = |d_1|$ we could accept $\mathcal{H}_o$ if and only if

$$\left| \sum_{i=1}^{n-1} A_i Y_i \right| \leq C,$$

where for given R_i, C is chosen from the relation

$$s\, P\left\{ \left| \hat{\delta} + \sum_{i=1}^{n-1} A_i Y_{oi} \right| > C \right\} = P\left\{ \left| \hat{\delta} + \sum_{i=1}^{n-1} A_i Y_{1i} \right| \leq C \right\} , \text{ with}$$

$$Y_{1i} = |d_1| \left(\frac{1}{R_{i+1}} - \frac{1}{R_i} \right) .$$

In order to compare the test of (42) with that of (15), we must compare the two test statistics

(44) (a) $$\sum_{i=1}^{n-1} \hat{A}_i \hat{Y}_i \equiv \hat{Q},$$

and

(b) $$\sum_{i=1}^{n-1} a_i \hat{Y}_i \equiv \hat{q},$$

and the two significance points

(45) (a) C_α (or C) and

(b) c_α (or c

if (15) is modified as was suggested above for (42)). We begin by noting that for the same reason as in (30).

$$\gamma = \begin{pmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_{n-1} \end{pmatrix} \equiv \text{a maximizes}$$

(46) $$h(\gamma) \equiv \sum_{i=1}^{n-1} \frac{\gamma_i \left(\frac{1}{r_{i+1}} - \frac{1}{r_i} \right)}{\sqrt{\sum_{j=1}^{n-1} \gamma_j^2 - \sum_{j=1}^{n-2} \gamma_j \gamma_{j+1}}}$$

while

$$\gamma = \begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_{n-1} \end{pmatrix} \equiv \text{A maximizes}$$

(47) $$H(\gamma) \equiv \sum_{i=1}^{n-1} \frac{\gamma_i \left(\frac{1}{R_{i+1}} - \frac{1}{R_i} \right)}{\sqrt{\sum_{j=1}^{n-1} \gamma_j^2 - \sum_{j=1}^{n-2} \gamma_j \gamma_{j+1}}} .$$

833478

We now assume that

$$\frac{\sigma_o}{r_i - r_{i+1}} << 1. \tag{48}$$

Recall that σ_o is the maximum possible error in range measurement (number 3 in section on Assumptions). We shall see in the next section that this assumption is satisfied. For example, if $\sigma_o = 50$ we may, in storing observations, only make use of those observations for which $r_i - r_{i+1} > 3700$. This already must occur in high-speed cases where relative speed is 1900 fps and one antenna revolution occurs every 4 seconds. Should antenna revolution rate be increased, in order for our results to be of much use, the error in range measurement would have to be decreased accordingly. The discarding of certain measurements is justified in the lower-speed cases because the suggested test gives such excellent results there even with the information being discarded.

We also assume that

$$\frac{\sigma_o}{r_i} << 1, \tag{49}$$

which naturally follows for $\sigma_o \leq 50$ due to the 12.5 seconds required warning time.

In order to compare the test statistics $\hat{Q}$ and $\hat{q}$, we take note of the following facts: $\hat{q}$ is a normal random variable with unit variance; given A_i, the value assumed by $\hat{A}_i$, $\hat{Q}$ is a normal random variable with unit variance. We shall show that no matter what value A_i is assumed by $\hat{A}_i$, $E\left(\hat{Q}|\hat{A}_i = A_i,\ i = 1, \ldots, n-1\right)$ is "close" to $E(\hat{q})$, and also that C_α and c_α are "close" to each other. This will prove that the test given by (42) and that given by (15) behave in almost the same way. In fact, the technique to be given will permit a quantitative comparison of the behavior of these two tests. In the development, we shall be guided by the approximations in (5).

Using these approximations, for given $r_1, \ldots, r_n$

$$E_d\left(\hat{Q}|\hat{A}_i = A_i,\ i = 1, \ldots, n-1\right) \doteq h(A)\ \frac{d}{\sqrt{2}\,\sigma} \tag{50}$$

and

$$E_d(\hat{q}) \doteq h(a)\ \frac{d}{\sqrt{2}\,\sigma}. \tag{51}$$

Our first task is therefore to show that $h(a)/h(A)$ is close to unity. This will be accomplished by first demonstrating that

$$\frac{h(a)}{H(A)} \text{ is close to unity,} \tag{52}$$

and then that

(53) $$\frac{H(A)}{h(A)} \text{ is close to unity.}$$

In proving (52) and (53), we shall make use of (48) and (49) as well as the statements concerning (46) and (47).

First we note, assuming $r_i > r_{i+1}$ (an assumption justified in the section on the sampling plan), from number 3 in the section on Assumptions, (48), (14), (41)

(54) $$\frac{1}{r_{i+1}} - \frac{1}{r_i} > 0, \quad \frac{1}{R_{i+1}} - \frac{1}{R_i} > 0, \quad a_i > 0, \quad A_i > 0.$$

It follows, using number 3 in the section on Assumptions, assuming $r_i > r_{i+1}$, that

(55) $$\frac{\frac{1}{R_{i+1}} - \frac{1}{R_i}}{\frac{1}{r_{i+1}} - \frac{1}{r_i}} < \left(1 + \frac{2\sigma_o}{r_i - r_{i+1}}\right) \frac{1}{\left(1 - \frac{2\sigma_o}{r_{i+1}}\right)}$$

and

(56) $$\frac{\frac{1}{r_{i+1}} - \frac{1}{r_i}}{\frac{1}{R_{i+1}} - \frac{1}{R_i}} < \left(1 + \frac{2\sigma_o}{R_i - R_{i+1}}\right) \frac{1}{\left(1 - \frac{2\sigma_o}{R_{i+1}}\right)}.$$

Let

(57) $$f = \min_i \left(r_{i-1} - r_i\right), \quad F = \min_i \left(R_{i-1} - R_i\right).$$

Then, using (54), we see that

(58) $$\frac{\left(1 - \frac{2\sigma_o}{r_n}\right)}{\left(1 + \frac{2\sigma_o}{f}\right)} \sum_{i=1}^{n-1} A_i \left(\frac{1}{R_{i+1}} - \frac{1}{R_i}\right) < \sum_{i=1}^{n-1} A_i \left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right)$$

and

(59) $$\frac{\left(1 - \frac{2\sigma_o}{R_n}\right)}{\left(1 + \frac{2\sigma_o}{F}\right)} \sum_{i=1}^{n-1} a_i \left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right) < \sum_{i=1}^{n-1} a_i \left(\frac{1}{R_{i+1}} - \frac{1}{R_i}\right).$$

53 3 1 7 8

Now, from (46), we know that

$$\sum_{i=1}^{n-1} A_i\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right) \leq \sum_{i=1}^{n-1} a_i\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right) \tag{60}$$

and from (47)

$$\sum_{i=1}^{n-1} a_i\left(\frac{1}{R_{i+1}} - \frac{1}{R_i}\right) \leq \sum_{i=1}^{n-1} A_i\left(\frac{1}{R_{i+1}} - \frac{1}{R_i}\right). \tag{61}$$

Combining (58) and (60), we have

$$\frac{\left(1 - \frac{2\sigma_o}{r_n}\right)}{\left(1 + \frac{2\sigma_o}{f}\right)} \sum_{i=1}^{n-1} A_i\left(\frac{1}{R_{i+1}} - \frac{1}{R_i}\right) < \sum_{i=1}^{n-1} a_i\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right), \tag{62}$$

while combining (59) and (61), we obtain

$$\sum_{i=1}^{n-1} a_i\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right) < \frac{\left(1 + \frac{2\sigma_o}{F}\right)}{\left(1 - \frac{2\sigma_o}{R_n}\right)} \sum_{i=1}^{n-1} A_i\left(\frac{1}{R_{i+1}} - \frac{1}{R_i}\right) \tag{63}$$

Equations (62) and (63) together with (48), (49) and (57) prove (52).

Using (56) and (57), we have

$$\sum_{i=1}^{n-1} A_i\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right) < \frac{\left(1 + \frac{2\sigma_o}{F}\right)}{\left(1 - \frac{2\sigma_o}{R_n}\right)} \sum_{i=1}^{n-1} A_i\left(\frac{1}{R_{i+1}} - \frac{1}{R_i}\right). \tag{64}$$

Equations (64) and (58) prove (53).

We have thus shown, combining (62) with (64) and (63) with (58), that

$$\frac{\left(1 - \frac{2\sigma_o}{r_n}\right)\left(1 - \frac{2\sigma_o}{R_n}\right)}{\left(1 + \frac{2\sigma_o}{f}\right)\left(1 + \frac{2\sigma_o}{F}\right)} \sum_{i=1}^{n-1} a_i\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right) < \sum_{i=1}^{n-1} A_i\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right)$$

$$< \frac{\left(1 + \frac{2\sigma_o}{F}\right)\left(1 + \frac{2\sigma_o}{f}\right)}{\left(1 - \frac{2\sigma_o}{R_n}\right)\left(1 - \frac{2\sigma_o}{r_n}\right)} \sum_{i=1}^{n-1} a_i\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right). \tag{65}$$

Note that for all practical purposes $R_n = r_n$, $F = f$. The term,

$$d \sum_{i=1}^{n-1} A_i\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right) = dh(A),$$

is within 0.42 percent of the expected value $E(\hat{Q} \mid \hat{A} = A)$, while

$$d \sum_{i=1}^{n-1} a_i\left(\frac{1}{r_{i+1}} - \frac{1}{r_i}\right) = dh(a)$$

is within 0.42 percent of $E(\hat{q})$, under (23) for $|d| \leq 1000$, as seen in (26).

Rather crude inequalities were used to obtain (65), and for any particular case the results could be sharpened greatly. We just sketch how this might be accomplished.

We note that

$$\sum_{i=1}^{n-1} a_i\left(\frac{1}{R_{i+1}} - \frac{1}{R_i}\right) = \frac{a_{n-1}}{R_n} - \frac{a_1}{R_1} + \sum_{i=2}^{n-1} \frac{a_{i-1} - a_i}{R_i}. \tag{66}$$

Thus, for given $r_1, \ldots, r_n$ (and therefore given a_i) we can compute, using (1), sharper upper and lower bounds for the left number. Thus, we can greatly improve (59) and thus (63).

Writing

$$A_i^* = -(n-i) \sum_{j=1}^{i} \frac{1}{R_j} + i \sum_{j=i+1}^{n} \frac{1}{R_j}, \tag{67}$$

we can get upper and lower bounds for A_i^*, and thus as above obtain upper and lower bounds for

$$\frac{\sum_{i=1}^{n-1} A_i^* \left(\frac{1}{r_{i+1}} - \frac{1}{r_i} \right)}{\sum_{i=1}^{n-1} A_i^* \left(\frac{1}{R_{i+1}} - \frac{1}{R_i} \right)} = \frac{\sum_{i=1}^{n-1} A_i \left(\frac{1}{r_{i+1}} - \frac{1}{r_i} \right)}{\sum_{i=1}^{n-1} A_i \left(\frac{1}{R_{i+1}} - \frac{1}{R_i} \right)}.$$

Thus, (58) can be improved, and therefore (62). Since this also improves (64), we have shown that (62), (63), and (64), (58) are improvable. Thus, (65) can be sharpened.

As an example of the results of the crude (65), if $r_n = 15{,}000$, $\sigma_o = 50$, $f = 3750$, then

$$\frac{\left(1 - \frac{2\sigma_o}{r_n}\right)\left(1 - \frac{2\sigma_o}{R_n}\right)}{\left(1 + \frac{2\sigma_o}{f}\right)\left(1 + \frac{2\sigma_o}{F}\right)} \doteq 0.94,$$

while if $r_n = 24{,}000$, $\sigma_o = 50$, $f = 7600$,

$$\frac{\left(1 - \frac{2\sigma_o}{r_n}\right)\left(1 - \frac{2\sigma_o}{R_n}\right)}{\left(1 + \frac{2\sigma_o}{f}\right)\left(1 + \frac{2\sigma_o}{F}\right)} \doteq 0.97.$$

To complete this section, we note that

$$\frac{\left(1 - \frac{2\sigma_o}{R_n}\right)}{\left(1 + \frac{2\sigma_o}{F}\right)} \sum_{j=1}^{n-1} A_j\, y'_{oj} \le \sum_{j=1}^{n-1} A_j\, Y_{oj} \le \frac{\left(1 + \frac{2\sigma_o}{f}\right)}{\left(1 - \frac{2\sigma_o}{r_n}\right)} \sum_{j=1}^{n-1} y'_{oj}, \tag{68}$$

where

$$y'_{oj} = \left|d_o\right| \left(\frac{1}{r_{j+1}} - \frac{1}{r_j} \right), \quad Y_{oj} = \left|d_o\right| \left(\frac{1}{R_{j+1}} - \frac{1}{R_j} \right).$$

When $d = \left|d_o\right|$, we see from (30) that with very little error we can take the expected value of

$$\sum_{j=1}^{n-1} A_j \hat{Y}_j$$

for given range measurements R_i to be

$$\sum_{j=1}^{n-1} A_j \, y'_{oj}.$$

Since

$$\sum_{j=1}^{n-1} A_j y'_{oj}$$

is proportional to $|d_o|$, the effect of using

$$\sum_{j=1}^{n-1} A_j Y_{oj}$$

rather than

$$\sum_{j=1}^{n-1} A_j \, y_{oj}$$

in computing C_α is to change the miss distance at which $P\{\text{acc } \mathcal{H}_o\} = 1 - \alpha$ by at most the percent indicated in (68). Thus, if $|d_o| = 200$, $\sigma_o = 50$, $R_n = 10{,}000$, $F = 5000$, the miss distance at which $P\{\text{acc } \mathcal{H}_o\} = 1 - \alpha$ is achieved is between 194 and 206 feet.

To summarize, we have shown the effect of replacing the test of (15) by a performable one to be small, giving quantitative results which estimate the "loss of vision." Equations (65) and (68) give the cruder general results. In the discussion following (68), a sharper method for estimating the "loss of vision" in any particular case, was outlined. In any case, the loss of vision can be made small by keeping error-in-range measurements suitably small (how small, to be found from (65), (68)).

THE SAMPLING PLAN–WHICH MEASUREMENTS TO STORE AND WHEN TO CEASE TAKING OBSERVATIONS

From the original assumption that all measurement errors, prior to discarding any observations, are mutually independent it can be shown that any sampling scheme for determining which measurements to use which is dependent on only the original observed range measurements, $R'_1, \ldots, R'_m$, will satisfy Assumption I. Due to space limitations, this will not be proved in this article.

Here is one reasonable sampling plan. Store R'_j, X'_j only if $R'_{j-1} > R'_j$. Estimate $|v|$ by $|v'| = (R'_1 - R'_s)/\tau_s$, where $R'_s - R'_1 > 13{,}000$, and τ_s is the time between measuring R'_1 and R'_s. (Note that since required warning range is at most 24,000 feet and detection

range at least 50,000 feet, this is feasible.) From $|v'|$ we can obtain a better estimate of the required warning range (error in $|v'|$ is approximately at most $2\sigma_o/\tau_s$ for small $|d|$, which is fairly small) and thus obtain near the maximum number of usable observations. Any subset of $\{R'_1, \ldots, R'_m\}$ for which successive observations are F feet apart will do for $R_1, \ldots, R_n$. This restriction is so that the results of the preceding section are applicable. As can be seen from an examination of test performance, it is desirable to let $R_1 = R'_1$, $R_n = R'_m$. We let $\hat{n}$ be the random variable whose value is the number of used observations, and $\hat{i}_1, \ldots, \hat{i}_{\hat{n}}$ be the random variables whose values denote the indices of those observations which are used.

COMPUTATION OF THE OVER-ALL PERFORMANCE OF THE TEST

We have seen that for any given r, x_1, $|v|$, we can compute upper and lower bounds for the probability of sounding an alarm. What is desired is to compute the probability of sounding an alarm for given $|v|$ and $|d|$ (or good upper and lower bounds). In theory, this is a simple matter, and is done as follows:

Give $r'_1, \ldots, r'_m$, x'_1, and the probability distribution of the R_i, from the sampling plan we can compute the probability distribution of $\hat{n}$, $\hat{i}_1, \ldots, \hat{i}_{\hat{n}}$. From this and the upper and lower bounds derived in an earlier section (replacing the Suggested Test by a Performable Test), we can compute upper and lower bounds for the probability of sounding an alarm for true ranges r', and first bearing angle x'_1, in the obvious way, i.e., we use the relation

$$P_{r'_1, \ldots, r'_m, x'_1}\{\text{sound alarm}\}$$

$$= \sum_{n, i_1, \ldots, i_n} P_{r', x'_1}\left\{\hat{n} = n, \hat{i}_1 = i_1, \ldots, \hat{i}_{\hat{n}} = i_n\right\} \cdot$$

$$\cdot P_{r', x'_1}\left\{\text{sound alarm} \mid \hat{n} = n, \hat{i}_1 = i_1, \ldots, \hat{i}_{\hat{n}} = i_n\right\}. \tag{69}$$

Now, given $|v|$ and $|d|$, we can assume some plausible probability distribution for the random variables m, $r'_1, \ldots, r'_m$, x'_1, and proceeding just as above compute upper and lower bounds for $P_{|v|, |d|}\{\text{sound an alarm}\}$.

This being rather laborious, we can obtain quick estimates of the performance of this test by looking at the "best" and "worst" cases.

For $|v| = 1900$ fps, assuming a detection range of 46,000 feet, antenna rotation time of $t_o = 4$ seconds, required warning time of 12.5 seconds, maximum error in range

measurement σ_o = 50 feet, and standard deviation of bearing-angle measurements σ = 0.00087 radians, roughly we have

Worst Case: $r_1 \doteq 46,500$ $r_2 \doteq 38,900$ $r_3 \doteq 31,300 = r_n$

Best Case : $r_1 = 46,800$ $r_2 = 39,200$ $r_3 = 31,600$ $r_4 = 24,000 = r_n$.

Setting $P\{\text{alarm}\} = 0.9995$ for $|d| = 200$ we find approximately:

d	P{alarm, worst case}	P{alarm, best case}
200	0.9995	0.9995
400	0.9454	0.5848
600	0.4616	0.0001
800	0.037	0.0000+ . . .
1000	0.0002	

REMARKS

In low-speed cases, where there is good reason to believe the constant velocity assumption seriously violated, it seems reasonable to require the system to warn the pilot when the second airplane is close, say closer than 5000 feet.

It is seen that an increase in bearing-angle measurement accuracy would improve the test's performance. Performance could also be improved if the number of observations per second were increased, while not permitting bearing-angle measurement accuracy to deteriorate seriously, and increasing the accuracy of range measurement correspondingly (see (65), (68)). This latter suggestion would seem the only feasible one in case of very high-speed aircraft, where with the given number of observations per second it might not even be possible to obtain two observations.

REFERENCES

[1] H. J. Landau and E. I. Pina, "Operational Requirements for a Collision Warning System," Operations Research, Vol. 5, No. 6, pp. 794-814.

[2] E. L. Lehmann, Testing Statistical Hypotheses, Wiley.

[3] R. Courant, Calculus, Vol. 1, Interscience.

* * *

LOGNORMAL DISTRIBUTION AND MAINTAINABILITY IN SUPPORT SYSTEMS RESEARCH*

R. L. Bovaird,† and H. I. Zagor‡

Technical Military Planning Operation - TEMPO
General Electric Company
Santa Barbara, California

This article presents a method for predicting the downtime distribution for a new complex electronic equipment. A theoretical discussion is given which indicates that the downtime distribution for such an equipment can be represented by a lognormal function. Experimental data are shown which establish that the two-parameter lognormal distribution function adequately describes the downtimes of a diverse group of electronic equipments. It is concluded that this analytical expression for the downtime distribution of complex electronic equipments provides the basis for a quantitative and scientific approach to the measurement and prediction of maintainability. This approach takes account of equipment maintenance characteristics and maintenance personnel capabilities.

INTRODUCTION

Of increasing importance is the recognition of maintainability as a major characteristic of military equipment. If maintainability is to have practical meaning in, for example, such areas as equipment specifications, it must have the qualities of being capable of definition, measurement and prediction. It is to this problem of measurement and prediction that we address this article, starting first with a discussion of downtime as a measurement of maintainability.[1]

DOWNTIME AS A MEASURE OF MAINTAINABILITY

A review of the literature dealing with the subject of maintainability shows that many different equipment characteristics and operational conditions are included under this heading.

---

*Manuscript received December 22, 1960.

†Now at Hughes Aircraft Company, Culver City, California.

‡Consultant to TEMPO; now at Radio Corporation of America, Defense Electronic Products, New York 13, New York.

The authors wish to express their gratitude to T. B. Slattery, Manager of the Logistics/Support Group at TEMPO for instituting the specific study which led to the final development of this paper; also to A. S. Goldman at TEMPO for his review and criticisms of the paper.

[1]For further discussion of the rationale which leads to this measure, c.f., Slattery, T. B., "Rationale for a Fundamental Measure of System Quantitative Definition of Maintainability", (RM 59TMP-72), Santa Barbara California: Technical Military Planning Operation, General Electric Company; 21 December 1959 [1] .

There are also many different viewpoints toward maintainability factors. As a result, a number of different measures and indices of maintainability have been found useful in one application or another, depending on which output functions and managerial decisions are of primary importance in each application.

For example, it can happen that the user of a system is interested in the fraction of demanded operation times that the system fails to perform satisfactorily. This is termed operational unavailability in this article and depends on the frequency of system failure (reliability), as well as on the duration of individual failures. In this case, an appropriate measure of maintainability is the downtime per failure which depends on both environmental parameters (e.g., the type of maintenance personnel, facilities, and spares available) and system parameters (e.g., fault detection and isolation features and ease of replacement of failed items).

To the user, maintainability is a function of the dollars spent for support operations. In the case just mentioned, at a given support cost the system with the least downtime per failure is the most maintainable. Also, the system with the lowest failure rate is the most reliable, and the system with the least total downtime in a given operational period is judged to be most supportable.

The system designer may not find these measures to be the most useful, for he is primarily interested in design parameters which he can influence, such as thoroughness of test, fault isolation capability, replacability of units in the system, and the like. For example, one design for a given system may permit 90 percent of its failures to be repaired in the field by isolating and replacing the faulty subassembly or part whereas an alternate design permits this for only 50 percent of its faults. The rest of system faults require unit replacement. To many designers, the second system is the more maintainable of the two. Others may judge the first to be more maintainable.

There are great difficulties in resolving conflicting views (such as those just mentioned), regarding an appropriate measure of maintainability acceptable to all. However, these difficulties must be overcome if satisfactory contractual agreements, regarding maintainability specifications, are to be written. In particular, an unequivocal measure of maintainability which can be measured and predicted will have to be found if this goal is to be accomplished.

One approach to this problem which appears promising is outlined in what follows. The suggested avenue of attack rests on the experimental observation that the distribution of downtimes per failure for a variety of electronic systems can be satisfactorily approximated by a lognormal distribution with zero mean. Further, the two parameters of this distribution appear to be correlated with the maintainability parameters of the system and of its environment.

The lognormal distribution provides a vehicle for quantitatively relating the various significant maintainability parameters in a form which is operationally meaningful. At the same time, guidelines are provided for structuring maintainability studies and discriminating between maintainability parameters, measures, and indices. Before indicating how this might be accomplished, the basic rationale and some supporting experimental evidence are presented.

CATEGORIES OF DOWNTIME

When it is observed that an equipment fails to perform its functions satisfactorily, all or part of it is taken out of operation to locate and correct the fault. The fault may be corrected by an adjustment or a part may be replaced with a spare. When it has been verified by appropriate test that the fault is corrected, the equipment is returned to service. It may be placed back in operation, or it may be placed in standby, depending on the operational conditions at that time.

Downtime can thus be divided into the following categories of time:

1. Recognition
2. Location
3. Correction
4. Verification.

In addition, part of downtime may be occupied with getting a spare or in waiting for tools or test equipment. The additional time required to do this, over and above the actual time required for the repair activities, is supply and administrative time. Further, the time between the actual occurrence of a fault in the equipment and its recognition is usually not known.[2] Downtime may, for most practical purposes, thus be divided as follows:

1. Locate time (time from recognition that a fault exists to its location)
2. Correction time (time from location of a fault to its correction, including post replacement adjustments, less supply time)
3. Supply time (time required to get a spare, over and above time required for actual repair activities)
4. Verify time (time from supposed correction of fault to verification by appropriate test).

It is to be noted that these categories of time refer primarily to a single functional group of equipment which is capable of only one mode of operation (unimoded). Whenever a failure of such equipment occurs, the equipment must be taken out of operation. If the equipment is multimoded it may not necessarily have to be taken out of operation; however, it can be analyzed by reducing it to a set of unimoded functional groups.[3]

DOWNTIME DISTRIBUTION

The total length of time an equipment is down for maintenance when a failure occurs varies statistically from one failure instance to another. This is also true for the previously noted individual time categories into which total downtime per failure is divided.

---

[2] In some systems this category of downtime contributes significantly to system unavailability, although it is extremely difficult to obtain quantitative data on the actual magnitude of these times. In such cases, one solution often adopted is to strive for maximum thoroughness of test in order to minimize this contribution to downtime.

[3] A system which can operate at several performance levels is multimoded. It can be divided into groups of equipment according to the specific functions performed during operation. The failure of some of its functional groups causes the system to pass into one of its alternate operational modes. (Cf. [4]).

Analysis of experimental data on downtimes shows that the observed range of variables can be fitted by a lognormal distribution. Figure 1 shows a typical fit of the downtime data by a lognormal distribution. The particular curve shown is for the APS-20E radar [2] but is representative of much of the data obtained [3]. A rationale which leads one to expect this result is discussed subsequently.

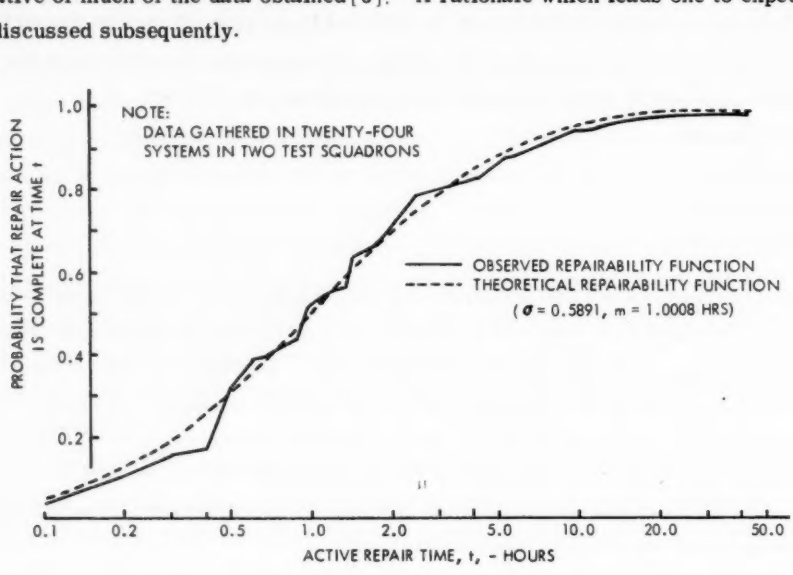


Figure 1 - Observed and theoretical repairability functions for AN/APS-20E radar systems

The lognormal distribution is one in which the logarithm of the distributed variable is normally distributed. The parameters which determine the function are the median value of the variable and the standard deviation of the logarithm of the variable. The frequency function can be written in the form

$$(1) \qquad f(t/m, \sigma)\, d(t/m) = \frac{1}{m\sigma\sqrt{2\pi}}(t/m)^{-1} \exp\left\{-\frac{1}{2\sigma^2} \ln^2 (t/m)\right\} dt,$$

where t = distributed variable,

m = median value of t,

and σ = standard deviation of ln t(dispersion parameter). [4]

The variable of the frequency function is seen in Eq. (1) to be t/m while σ^2 appears in the exponential factor. The parameter σ thus determines the shape of the function while the parameter m determines the scale on its variable. The multiplying constant $1/(m\sigma\sqrt{2\pi})$ normalizes the frequency function so that its integral over the entire range of t/m equals unity. This

[4] The density function of a variable x, $0 < x < \infty$, where ln x is normally distributed with mean μ and standard deviation σ, is

$$f(x, \mu, \sigma)\, dx = (1/\sigma x\sqrt{2\pi}) \exp\left[-(\ln x - \mu)^2/2\sigma^2\right] dx.$$

If $\mu = 0$ and $x = t/m$, this reduces to Eq. (1). Note that m may be any constant or parameter of proper dimension. It is convenient here to let m be the median value of t.

frequency function is plotted in Figure 2 for several values of σ. The corresponding distribution functions are plotted in Figure 3. The lognormal frequency function (Figure 2) is seen to be asymmetrical with a long tail to the right and a sharp cutoff to the left. As the dispersion parameter σ becomes smaller, the distribution becomes more symmetrical. It is seen in Figure 3 that the distribution is very nearly symmetrical and normal for values of σ less than about 0.2. The limit as $\sigma \longrightarrow 0$ is a normal distribution function with a standard deviation of zero.

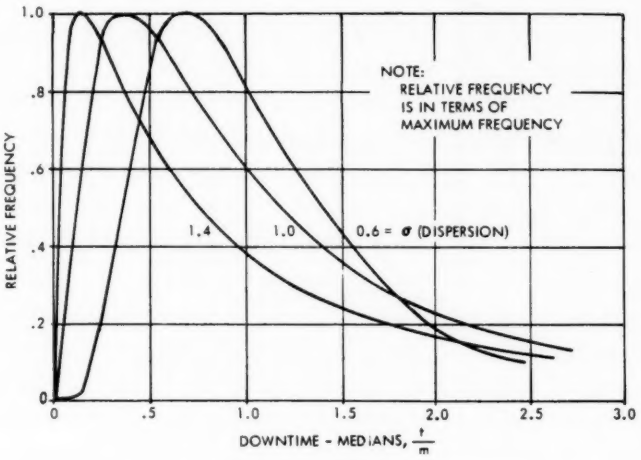


Figure 2 - Lognormal frequency function

As σ (Figure 3) increases, the distribution of observed values of the variable is dispersed over a larger range although the median or half way value remains constant. The small values to the left of the median are relatively evenly distributed from 0 to m, while the large values (i.e., those which lie to the right of the median), are distributed relatively evenly out to very large values (i.e., the tail of the frequency function becomes very long and uniform).

It can be shown [5] that the arithmetic mean "$\bar{t}$" (expected value of the variable) is given by

(2) $$\bar{t} = m\, e^{\sigma^2/2},$$

where m = median,

e = base of natural logarithms (e = 2.718),

and σ = dispersion parameter.

Inspection of Eq. (2) shows that for a given σ the expected or average value of the variable is directly proportional to the median. Equation (2) also shows that, for a fixed median, the average value increases exponentially with the square of the dispersion parameter. For this reason it is important in a practical case to keep σ as small as possible.

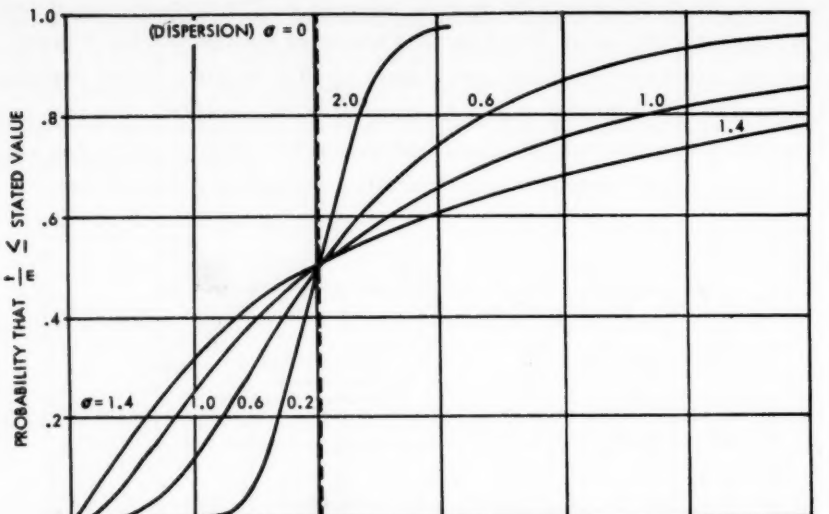


Figure 3 - Lognormal distribution function

Lognormal Theory

The lognormal distribution function is frequently observed in two general areas of phenomena. One of these is in connection with growing organisms; the other is in connection with partitioning processes (e.g., placing objects in logically derived categories). There is as yet no general theory which gives the necessary and sufficient conditions for the occurrence of a lognormal distribution. However, it has been shown that whenever the "law of proportionate effect" rules a statistical variable, it is lognormally distributed [5]. The law of proportionate effect holds whenever the change in a variable at any step of a process is a random proportion of the previous value of the variable.

It has been known for a long time that response times of animals and humans to simple physical stimuli or perceptions are frequently lognormal (the so-called Weber-Fechner law). More recently, experimental studies of the reaction time of humans to more complicated perceptual patterns involving some degree of learning indicate that these reaction times are also frequently lognormal.

It has also been observed that "when a large number of items is classified on some homogeneity principle, the variate defined as the number of items in a class is often approximately lognormal. Examples of this phenomena...are the number of persons in a census population class, the number of Sino-Japanese characters in a lexicographical group, and the outlay by households on classes of commodities,"[5]. Another reported observation is that the number of spores per leaf of a tree is lognormally distributed [5]. According to the above statement regarding classification on a homogeneity principle, this fact is a consequence of the manner of counting spores rather than an attribute of the spores on a tree.

Horvath has noted that the service times at a tool crib and in a library can be fit very well by a lognormal distribution [6]. The process of locating a demanded tool or book is a partition process characterized by logic, for a coordinate system is used to find the tool or book (aisle, shelf, bin or place on shelf). In both cases, a large number of items is classified on a homogeneity principle (location). This is also true to a certain extent for faults in an electronic equipment, for a fault is located to a certain level, say to within a component or module.

These observations suggest very strongly that the human partition process of placing objects in logical categories somehow leads to a lognormal distribution of the number of objects in a category. The precise manner in which this takes place and the necessary and sufficient conditions for a logical process to produce a lognormal distribution have not been discovered. However, these considerations indicate that it is not surprising that maintenance downtimes for electronic systems are lognormally distributed, since location and correction of a fault in such an equipment is to a large extent a logical process of placing things in categories. The categories are functional assemblies or components, and failures are categorized according to the failed component.

Practical considerations indicate that the distribution of downtimes should be asymmetrical. As a rule, most of the maintenance personnel are familiar with the operation and repair of the most frequently occurring malfunctions in systems that have been in service a number of years. At relatively infrequent intervals, however, an unfamiliar trouble occurs or an inexperienced technician is called on to repair the system. Such repairs require a longer time than most repairs. This type of maintenance situation exhibits a pattern of repair times which cluster about a relatively low value, but in which the distribution shows a tail to the right. Thus it is unimoded and skewed to the right. One member of a class of such curves, the lognormal distribution, describes this situation very well over the observed range of variate (i.e., downtimes).

It is to be noted that other distributions of this class (e.g., Pearson Type III or Chi Square) may describe the observed data equally well. However, these distributions do not arise from the partitioning process which is inherent in the maintenance activity.

An interesting property of the downtime data was observed by Howard [7]. The lognormal distribution function fitting a particular sample of data was found by fitting a straight line to the plot against downtime on lognormal probability paper of the cumulative fraction of the downtimes not exceeding a given magnitude. These same data were also plotted on semilog paper and a curved line was obtained, showing that this function is not exponential. Were this curve a straight line, the function would be exponential. By fitting a straight line to the tail of this curve and removing from the sample the downtimes falling in this exponentially fitted portion, a smaller sample was obtained (the remainder) for which the process was repeated. In this way, Howard [7] was able to approximate the lognormal function by a sum of exponential functions. Further, it was found that the distribution function remaining after each removal of an exponential function was lognormal, although it approached an exponential function as more exponential distributions were removed from the sample. In general, one would expect the sum of exponentials to give a Pearson Type III distribution [8]. This indicates that over the observed range of variable, the data can be equally well fitted by the Pearson III (i.e., Chi Square) distribution.

Howard [7] also noted that the downtimes which were grouped together by the above process fell more or less into definite physical categories (e.g., routine adjustments, normal replacements of frequently failing part, and the like). The downtime distribution within each of these categories was, to a very good approximation, exponential, while the distribution obtained by placing them together was lognormal.

There is a growing body of evidence which indicates that, in general, downtime distributions within the various categories are either lognormal or exponential; and the total downtime distribution is lognormal, provided the equipment is sufficiently complex. Otherwise it tends to be exponential.

Experimental Data

Observed downtime distributions for a number of equipments have been reported. Howard [7] has plotted total downtime for the fire control system for two large surface-to-air-missiles and also total downtimes for a heavy surface radar. Horne [9] plotted active work times in the aircraft and in the repair shop for the AN/ARC-27 communication sets of two different aircraft squadrons and also active work time in the aircraft for the AN/APS-20E radar. Scott [3] plotted total downtime distributions for three surface shipboard equipments (AN/SRR-13A Radio Receiver, AN/SPA-8A Radar Repeater, AN/SPG-48FCS) and listed the distribution determining parameters for two other equipments (AN/URT-2 Radio Transmitter, TED-4 Radio Transmitter). Scott also showed the distribution of part procurement time, preparation time, fault location time, and fix time. In these data the fault location time, fix time, and total downtime always showed a lognormal distribution, but the preparation time and part procurement time were not as consistent. Table 1 lists the dispersion parameter σ, the median downtime m, and the average downtime for all the equipments for which data were available.

It may be seen in Table 1 that the data fall into two groups according to whether σ has a low or high value. The first group includes the three distributions of repair times obtained on board aircraft. The second group includes all the time distributions obtained on board ships and on the ground. The average σ for the first group is 0.60 while that for the second group is 1.40.

The difference between these two groups lies in the type of maintenance performed. Airborne equipment like the APS-20E and ARC-27 are primarily maintained in operating condition by means of module replacement, whereas ground equipment maintenance may also include the repair of faulty components and part replacements, rather than just module replacements. Also, the priority of ground maintenance is not usually as great as is that of airborne maintenance, for there may be spare systems available on the ground for immediate operation. For relatively simple equipments like radios this is usually the case. The situation is different, however, for a complex system like a surface-to-air missile fire control system.

It is to be noticed that the TED-4 transmitter has a $\sigma = 0.69$ and $m = 0.561$ while the URT-2 transmitter has a $\sigma = 1.41$ and $m = 1.21$. These are similar equipments even though the σ's and m's are quite different. It turns out that the TED-4 is manually tuned and a relatively simple equipment while the URT-2 is pushbutton tuned and relatively complex. The latter gives a great deal of maintenance trouble.

The SPA-8A is a highly modularized system in which a failed tube, for example, is replaced by replacing the module containing the tube. This would lead one to expect that it should be in the upper group in Table 1. However, the equipment has an ineffective fault location system (locating the failed module is difficult). The maximum maintainability permitted by its modular construction is thus not obtained.

TABLE 1
Experimental Data on Lognormal Distribution

System	σ	m	$\bar{t}$	Time Category	Place of Repair
ARC-27 Communications Set	0.46	0.483	0.537	Active repair time	Aircraft, Group B
APS-20E Radar	0.589	1.008	1.196	Active repair time	Aircraft, Group B
ARC-27 Communications Set	0.646	0.507	0.625	Active repair time	Aircraft, Group A
TED-4 Transmitter	0.69	0.561	0.710	Total downtime	Shipboard
ARC-27 Communications Set	1.08	1.83	3.28	Active repair time	Ground, Squad. B
SPA-8A Radar Repeater	1.12	0.809	1.51	Total repair time	Shipboard
FCS-SAM	1.16	0.65	1.27	Total downtime	Ground
ARC-27 Communications Set	1.17	1.84	3.65	Active repair time	Ground, Squad. A
Heavy Radar	1.39	4.0	10.57	Total downtime	Ground
URT-2 Transmitter	1.41	1.21	3.26	Total downtime	Shipboard
SRR-13A Receiver	1.50	1.08	3.32	Total downtime	Shipboard
SPG-48 FCS	1.76	1.01	4.75	Total downtime	Shipboard

Heuristic Considerations

The relatively small variation in σ among the diverse electronic equipments (Table 1) can be understood when it is noted that electronic equipments all use similar types of components and that the manner in which these components are connected and arranged to perform a function is generally similar, regardless of the function. Further, a complex equipment can be viewed as being a logically connected set of simple equipments or functional groups. The designers of such equipments use very similar engineering techniques and principles in designing the fault location system. Usually, the initial fault recognition includes an identification of the failed functional group. As a result, greater system complexity due to an increase in the number, but not complexity, of functional groups comprising the system does not generally increase the time to locate and correct a fault, but it does

increase the failure rate of the equipment. On the other hand, greater system complexity due to an increase in the internal complexity of the distinct functional groups can appreciably increase the time to locate and correct a fault as well as increase the failure rate.

This can also be viewed from the human factors aspect. In general, military maintenance crews will exhibit similar characteristics in that each is composed of people with similar training in similar environments and required to do similar kinds of tasks. In addition, the individual members of the maintenance crews are subjected to a renewal process in the natural personnel turnover in the Armed Services. Thus, in a statistical sense after a period of time, one would expect the various crews who deal with similar types of equipments to yield similar maintenance time distributions. However, the median time to perform maintenance is not expected to be a fixed quantity. These differences in the median values, m, arise in that some equipments are inherently more difficult to maintain than others, even though, on the average, similar techniques of application and tasks are required. Thus the maintenance personnel's qualitative performance will be similar, but the actual times required to do these tasks will vary between the equipments.

Another factor is personnel motivation or incentive. In an aircraft, there may be more urgency to effect maintenance. Thus, one would expect the medians to be different from those obtained on aircraft maintenance action as compared with those same equipments when they are having ground shop maintenance work done on them. In the case of the ARC-27 data, the dispersion parameter, σ, as well as the median, m, are different for aircraft and ground repair shop maintenance.

The line of reasoning just considered leads to the conclusion that variations in σ and m among electronic equipments indicate variations in approach to the design of failure detection systems, personnel skill levels, motivation, troubleshooting routines, and component replacement techniques. Thus, variations in σ and m among various electronic systems do not usually indicate variations in type or complexity of the comprising equipments, except as these factors affect the engineer's ability to design good fault location systems and to arrange components physically so that they are easy to replace.

The dispersion parameter, σ, and the median m, of the lognormal distribution function describing downtimes of an electronic system are indicators which are sensitive to the maintainability of the system. They are directly determined by the interaction between the maintenance characteristics of the system and the maintenance skill and motivation of the maintenance crew in the given environment. The parameter, σ, indicates the relative proportion of short, medium, and long downtimes (i.e., the variability in maintenance performance), while the parameter, m, indicates the general level or magnitude of downtimes experienced (i.e., the level of maintenance performance).

PREDICTING DOWNTIMES

The preceding data and discussion indicate a way to predict downtimes for an electronics system which has but one mode of operation. Although there are reasons for believing that multimoded system downtimes also follow the same distribution function, the present discussion

is restricted to unimoded systems. A unimoded system is either nonfailed in its entirety, or else it is failed and out of operation. The method for predicting the downtime distribution for such a system rests on the fact that the form of the distribution function is known to be closely approximated by a lognormal function. This function is completely determined by the two independent parameters σ and m. These parameters were observed to be strongly correlated with maintenance factors which can be estimated for a given equipment. Knowledge of the way in which σ and m vary with the factors which determine them is as yet fragmentary. This knowledge can be increased, nevertheless, by analyses of downtime data and from experiments using equipment and personnel under controlled conditions designed to provide data on specific equipment.

Examination of the data shown in Table 1 and the curves of Figure 3 shows that $\sigma = 1.4$ is a value for the dispersion parameter which is approximately correct for a wide range of electronic systems. This value is applicable to a system when its maintenance includes fault location, equipment adjustments, and spares replacement at all equipment levels from individual parts to full assemblies of parts, components, subassemblies, and replaceable modules.

Another wide range of electronics systems is approximately described by $\sigma = 0.6$. This second class of systems is repaired primarily by module replacement, and fault location to the module level is a relatively simple process. Further, administrative and spares procurement delays are virtually eliminated in this type of maintenance as, for example, in inflight maintenance of airborne systems during the time the equipment is required to operate.

Estimation of the appropriate median downtime for a particular system is more difficult, as inspection of Table 1 shows. However, in the authors' opinion it would be relatively easy for experienced maintenance personnel to estimate this parameter approximately for a specific system, provided the expected operational maintenance environmental conditions are quite specifically described. The variation of this parameter with the various factors which determine it can be discovered in any particular case through appropriate experimental measurements.

POSSIBLE APPLICATIONS

The observations of the preceding sections show that it is possible to study maintainability in a quantitative way. An approach has been indicated. The approach places the measurement and prediction of equipment downtime at the center of the study of maintainability and relates both system parameters and operational parameters to this maintainability measure. This measure in turn is given operational significance through its effect on system availability for satisfactory functional performance. Downtime thus provides a means of deriving quantitative functional relationships between system characteristics, the operational environment, and significant operational effects. A way by which this might be accomplished is now outlined.

Maintainability and Availability

The user of a system is interested in its operational availability (i.e., in the fraction of total demand time that the system is capable of performing required functions satisfactorily). This important operational measure of support effectiveness depends on the nature of user

demands, on system characteristics, and on the capabilities of the maintenance and logistic support organizations. These combine to yield system operational availability in the way shown in Figure 4.[5]

Figure 4 shows that from the user's point of view, maintainability has two major elements: maintenance and logistics. Maintenance refers to fault location and removal activities whereas logistics refers to the supplying of material, particularly spare items required to support the maintenance activity. The time required for each of these two activities determines the time to return the system to service once it has failed. This is the duration of a failure which, together with the frequency of failure, determines the total downtime (and hence uptime) experienced by the system during a period of operational use.

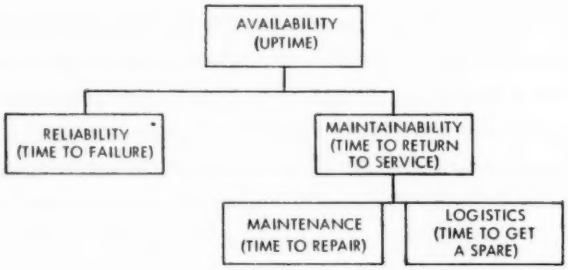


Figure 4 - Relation of maintainability elements to availability

Maintainability Parameters

The time to repair a failure (excluding supply time) depends strongly on the operational conditions at the place of repair (such as the number and capabilities of the maintenance personnel); the tools, test equipment and manuals available to them and the nature of the maintenance facilities. The time to get a spare depends on the distance of the supply stores from the place of repair, on the administrative procedures required to obtain a spare, and on the nature of the logistics communication and transportation facilities. In addition to these operational factors, repair and supply times are also strongly influenced by system characteristics such as built-in fault detection and location capabilities, and equipment packaging and installation for removal and placement of failed items.

Parameter Investigation

The system designer can influence system characteristics affecting support effectiveness whereas the system user can influence operational conditions important for system sup-

[5] The point of view in Figure 4 is developed in greater detail in [10]. See also [11] for a closely related treatment of this subject.

port. In addition to support parameters which can be controlled, there are many others which are uncontrollable and frequently random in nature. This constellation of support parameters determine the operational availability of the system by way of operational reliability and maintainability. The approach to maintainability measurement described in the previous sections of this paper provides a means of investigating quantitatively the functional relationship between operational availability and the many parameters which determine it. Figure 4 shows a conceptual framework in which this might be done which relates the user parameters to those of the designer.

One avenue of investigation which appears rewarding is a study to determine the precise manner in which the two parameters of the lognormal distribution (σ and m) are related to the maintenance factors which determine them. The result would permit accurate prediction of the expected downtime distribution for a new system and would provide a means of measuring and controlling the operational maintenance performance of an equipment and its maintenance crew. Of particular importance in such a study is the discovery and evaluation of the precise effect on maintenance performance of various human factors such as training, experience, motivation, and operational environment.

REFERENCES

[1] T. B. Slattery, Rationale for a Quantitative Definition of Maintainability (RM 59TMP-72), Santa Barbara, California: Technical Military Planning Operation, General Electric Company, December 1959.

[2] R. C. Horne, AN/APS-20 E Radar System Maintainability Study (Status Report No. 1, Project 0315, Contract NObsr-7757), Washington, D. C.: ARINC Research Corporation, 15 July 1959.

[3] R. C. Horne and S. R. Scott, BuShips Maintainability Study (Interim Report No. 1, Contract NObsr-77557), Washington, D. C.: ARINC Research Corporation, 21 August 1959.

[4] R. Bovaird and H. I. Zagor, A Systems Approach to Specifying and Measuring Polaris FCS Operational Availability (RM 59TMP-57), Santa Barbara, California: Technical Military Planning Operation, General Electric Company, December 1959.

[5] J. Aitchison and J. Brown, The Lognormal Distribution, Cambridge University Press, 1957.

[6] W. J. Horvath, Jr., Applicability of the Lognormal Distribution to Servicing Times in Congestion Problems, Operations Research Society, January-February 1959.

[7] F. A. Hadden, R. R. Howard, and W. J. Howard, Study of Downtime in Military Equipment, Proceedings 5th Annual Symposium on Reliability and Quality Control, IRE-EIA-ASQC, January 1959, p. 402.

[8] Harold Cramer, Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946.

[9] R. C. Horne, Investigation of the Effects of Different Maintenance Procedures on the Reliability and On the Maintainability of AN/ARC-27 Communications Sets (Status Report No. 1 Project 0314), Washington, D. C.: ARINC Research Corporation, 1 June 1959.

[10] R. L. Bovaird, A. S. Goldman, and T. B. Slattery, Concepts in Operational Support Research (RM 60TMP-70), Santa Barbara, California: Technical Military Planning Operation, General Electric Company, November 1960.

[11] E. L. Welker and R. C. Horne, Jr., Concepts Associated with System Effectiveness, Monograph No. 9 ARINC Research Corporation, Washington, D. C., 15 July 1960.

* * *

THE MINIMAX PATH IN A SEARCH FOR A CIRCLE IN A PLANE*

Brian Gluss

Armour Research Foundation of Illinois Institute of Technology

The problem has been considered by Isbell of determining the path that minimizes the maximum distance along the path to a line in the same plane whose distance from the starting point of the search is known, while its direction is unknown. We consider in this article the analogous problem for the search for a circle of known radius, of known distance from a starting point in its plane. It is further shown that Isbell's solution is a limiting case of the problem posed as the radius of the circle tends to infinity. The problem appears to be of some practical significance, since it is equivalent to that of searching for an object a given distance away which will be spotted when we get sufficiently close—that is, within a specific radius.

INTRODUCTION

Suppose that it is desired to trace a path from a point A (Figure 1) in such a way that it minimizes the maximum distance travelled along the path until a coplanar circle, with center O' and radius s, is reached, given only that the distance from A to O' is r+s.

A heuristic argument will be presented justifying the solution that is given below, although a rigorous proof could be demonstrated, which however would be lengthy and tedious. We shall present a couple of examples for different values of r/s, and shall also demonstrate that as $s \to \infty$ the solution tends to Isbell's [1] optimal path for hitting a straight line a fixed distance r away.

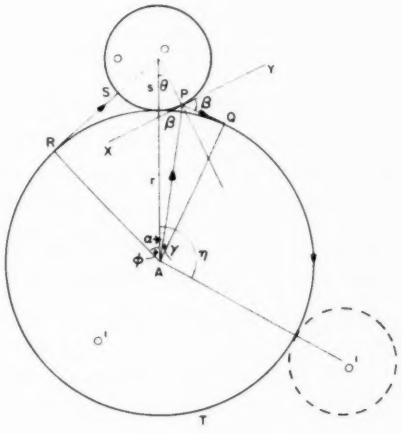


Figure 1

THE SOLUTION

The path comprises the straight lines AP, PQ, the arc QTR of the circle ⊙', center A, radius r, and the straight line RS, where the points P, Q, R, S and the lines satisfy the following properties. Let OA be an arbitrary straight line of length r + s and let ⊙ be the circle with center O and radius s. Then P is on ⊙; PQ is a tangent to ⊙'; also, P and Q

*Manuscript received August 9, 1961.

are such that the tangent XPY to ⊙ at P has the property that the angles $X\hat{P}A$, $Q\hat{P}Y$ are equal; finally R and S are such that the tangent at R to the circle ⊙' passes through S, which is on ⊙ and through O.

Firstly, it should be noted that, if $\eta = O\hat{A}O'$ is the angle that the line joining A to the center of the circle it is desired to reach subtends to AO, then (a) the path APQ covers the possibilities $0 \leq \eta \leq O\hat{A}Q$, (b) the arc QTR covers the cases $O\hat{A}Q < \eta \leq O\hat{A}R$, where $O\hat{A}R$ π and (c) RS accounts for $O\hat{A}R < \eta \leq 2\pi$.

Let us first justify the path RS. After leaving ⊙', it is necessary to hit ⊙ in order to cover the remaining possibilities for $\eta \leq 2\pi$. It then follows that the path of shortest distance that leaves ⊙' at a point R and hits ⊙ at S is such that RS is the tangent to ⊙' that passes through O. This is because (a) if we leave ⊙' before the prescribed R, then we must remain outside it in order not to miss possible positions of the circle we wish to find that would be missed if we went inside ⊙', and then with this restriction the path is obviously greater than the one prescribed, (b) if we leave the circle after R all possible paths are greater than RS, and (c) if we leave at R, RS is the shortest path to ⊙.

Now consider the path APQ. Let us first show that, of all paths hitting ⊙ at P' and then proceeding to Q' on the circle ⊙', the one prescribed is best. Firstly, to minimize AP'Q', given Q', AP' and P'Q' must be straight lines. Now the ellipse through P' with foci A and Q' is such that $S = AK + KQ'$ is a constant S' for K on the ellipse, $S < S'$ for K within the ellipse, and $S > S'$ for K outside it. Hence it follows that the ellipse with foci A and Q' that touches ⊙ minimizes $AP' + P'Q'$ for fixed Q'. If this ellipse touches at $P' = P''$, the well-known elliptic property that the angles the tangent subtends to the foci are equal implies that the angles $X'\hat{P}''A$, $Y'\hat{P}''Q$ are equal, where $X'Y'$ is the tangent at P'' common to ⊙ and the ellipse. The optimal path also has the property that $P''Q'$ must be a tangent to the circle ⊙' Hence, the only path having these properties is that for which P'' is P and Q' is Q.

This completes the heuristic proof. It should be noted that, in effect, we have restricted ourselves to convex paths.

EQUATIONS FOR COMPUTATION OF SOLUTIONS

From Figure 1 it may be seen that the following equations are satisfied by the solution.

(1) $$2\beta - \gamma = \frac{\pi}{2},$$

(2) $$\alpha + \beta = \frac{\pi}{2} - \theta,$$

(3) $$AP = r \sec \gamma,$$

and, from the triangle AOP,

(4) $$\frac{\sin \alpha}{s} = \frac{\sin (\beta - \gamma)}{r + s} = \frac{\sin \theta}{r \sec \gamma}.$$

Manipulation of these equations to eliminate α, γ and θ, finally gives the equation for β:

$$\frac{r}{s}(1 - \sin^2 2\beta) = 4 \sin^2 \beta \cos \beta \, (2 \cos \beta - 1), \tag{5}$$

from which β may be easily computed for particular values of r and s by successive approximation.

LIMITING CASES

1. $s/r \to 0$

As $s/r \to 0$, it is obvious that the solution should tend to a path in which we proceed along the radius to the circle $\odot'$ until the perimeter is reached, and then go round the circle. Equations (1) to (4) confirm this, in that α, γ and $\theta \to 0$. It might incidentally be noted that $\beta \to \pi/4$.

2. $s \to \infty$

As $s \to \infty$, the problem tends to Isbell's problem of searching for a straight line. It should first be noted that ϕ satisfies the equation

$$\cos \phi = \frac{r}{r+s}. \tag{6}$$

Then, $\phi \to \pi/2$ as $s \to \infty$, and from Eq. (5), $2 \cos \beta - 1 \to 0$ or $\beta \to \pi/3$, since the other solutions $\beta = 0, \pi/2$, to Eq. (5) obviously do not give meaningful paths. The values $(\beta, \phi) = (\pi/3, \pi/2)$ correspond to Isbell's solution.

TWO EXAMPLES

Before computing the solution for two particular values of r/s, we should first give the equation for the maximum distance D travelled along the path; that is, the distance A P Q T R S:

$$D = r[\sec \gamma + \tan \gamma + 2\pi - \alpha - \gamma - \phi + \tan \phi] - s. \tag{7}$$

1. $r/s = 1$

Solving the Eq. (1) to (7), we obtain

$\phi = 60°$,
$\beta = 56° \, 36'$,
$\gamma = 23° \, 12'$,
$\alpha = 15° \, 59'$,
and $D = 6.8007s$.

2. $r/s = 3$

$\phi = 41° \, 25'$,
$\beta = 53° \, 50'$,
$\gamma = 17° \, 40'$,
$\alpha = 8° \, 29'$,
and $D = 21.0721s$.

It can incidentally be seen that the distances D are a little smaller than $r + 2\pi r$, 7.2832s and 21.8450s respectively, the distances corresponding to paths proceeding radially to the circle $\odot'$ and then around its perimeter.

ACKNOWLEDGMENT

The author is indebted to Mr. Robert W. Floyd for helpful discussions on the subject.

REFERENCE

[1] J.R. Isbell, "An Cptimal Search Pattern, "Naval Research Logistics Quarterly, Vol. 4, No. 4, December, 1957, pp. 357-9.

* * *

MATERIAL LOGISTIC SUPPORT OF WEAPON SYSTEMS*

D. W. Whelan, SC, USN

BACKGROUND

In recent years the experts on defense posture have spoken with ever increasing frequency on the subject of weapon systems. Neophytes on the subject have been at once charmed and confused by the term. It would appear that some of the experts participate in the confusion if the vagueness with which they sometimes describe a weapon system is any indication of the exactness of the definition.

The evolution of Naval warfare has been from massive fire power toward sophisticated weapons which have the capability singly or in small salvos to seek out and destroy the target by increasing the kill probability of each warhead. The evolution is marked by the transition from thousands of rounds of heavy ammunition to hundreds of surface launched guided missiles; from millions of rounds of aircraft ammunition and tons of bombs to thousands of air launched missiles.

The World War II concept of gaining the objective through saturation by volume firepower has yielded to highly selective placement of highly accurate missiles. This is the era of measuring distances in miles and speed in mach numbers. This is the era of measuring in seconds the time to acquire targets, select warheads, firing order, firing, and reloading to repeat the cycle. Accomplishing this feat presumes high-speed electronic computation of the fire-control problem and presupposes the earlier imput of a myriad of instructions to the computer.

These are the elements of the weapon system. CNO defines a Navy Weapon System as the combination of a weapon or multiple of weapons and the equipment employed to bring the destruction of the weapon against the enemy. Navy weapon systems may be shipborne by surface ships or submarines or they may be airborne. By definition, within a Navy weapon system the weapon itself is associated with the detection, location and identification devices necessary for the employment of the weapon; the delivery units employed to launch the weapon and the control units used to exercise control of the weapon after it has been launched.

DEFINING WEAPONS SYSTEMS

It is vital to any consideration of the support of the weapon system of tomorrow's Navy to define definitely and precisely the equipments and the personnel who make up the weapon system. It is not necessary to contrive a single definition of a weapon system into which all weapons fit nicely. It appears best for the purposes of the Navy, to define each weapon system by listing all the related equipments which are required for the full and proper functioning of the weapon. Any definition in these terms requires some arbitrary decisions

*Manuscript received January 1960.

as to the extent of the weapon system in order to preclude the possibility of the definition becoming so monstrous that it defies handling by a closely enough knit organization. For example, a ship's electric, pneumatic, and hydraulic systems may support a number of weapons as well as administrative and housekeeping functions in the ship. The service-type facilities are better considered systems in themselves and are better supported as such. But, again, each weapon system must be clearly defined before any specific consideration can be given to the material support of the weapon system.

WEAPON SYSTEMS SUPPORT ORGANIZATIONS

When the decision was made to develop an intermediate-range ballistic missile (IRBM) capability for the Navy on a compressed time schedule, the task was assigned to the Bureau of Ordnance Special Projects Office. A number of facts and proposed goals precipitated this decision. The benefit of such an organization lies in the fact that the organization has a single purpose and the authority, staff, and funds for its realization. The task itself is ambitious enough to stagger the layman's imagination but the organization which was created for the purpose of accomplishing it provides for the most effective coordination of effort for its realization.

Other missile systems such as Regulus and Terrier have been developed and made operational within the regular basic organizations of the Bureau of Ships and the Bureau of Ordnance. That these have been noteworthy accomplishments, goes without saying. But that there appears to be a more efficient method for the accomplishment of desired goals in new weapon systems, is evidenced by the recommendation of the committee on organization of the Navy that BUORD and BUAER be consolidated in a Bureau of Naval Weapons. The organizational pattern of the Bureau of Naval Weapons considers under the Assistant Chief for Program Management the vertical organization requirements of the weapon systems. It follows the old organization axiom that where two or more people or groups of people, at the same level, labor on separate subtasks which are contributing to the whole task, there must be a common supervisor or supervisory group that provides the central direction for established projects.

Just as the term weapon system has been the victim of several definitions, the concepts of logistic support organizations on a weapon system basis have been described and employed in several ways.

It has, for example, been the practice of the U. S. Air Force— when a missile system graduates from the Research and Development stage to production—to assign logistic responsibility to one of the several air material areas. These air material areas are designated as lead air material areas (LAMA's) for the weapon systems assigned to them for program support. These activities become, in effect, a combined technical bureau and SDCP for their weapon systems. It is inherent in the concept of vertically integrated systems of logistic support for each weapon system that there will be built in plural managers of many common items of supply and a certain amount of material cross hauling. Such a logistic system has built into it duplication among weapon systems on a functional assignment basis and material item basis.

In the Navy we consider this a disadvantage when it is contrasted with the comparatively clean assignments for material management among inventory managers by cognizance assignment and separation of functions between technical bureaus and program supply demand control points. The apparent disadvantage to the Air Force, however, must be evaluated in the light of the offsetting advantage which centralizes logistic support responsibility for each weapon system in the hands of a clearly identified logistic manager (LAMA).

The debate between vertical weapon system support management and an integrated supply system may be illustrated by the example of a recent provisioning of an FBM equipment. The provisioning decisions developed a list of repair parts in which some $900,000 would be assigned for inventory management of ordnance parts and about $50,000 for common electronic and general stores material. No actual tally was kept on the number of man hours required to accomplish the ordnance parts procurement and supply support required. However, it is estimated conservatively that the effort in the provisioning activity per item to obtain supply support was twice that required to procure and establish each ordnance item. This observation is not intended as criticism of any inventory control point. It is intended only to illustrate the point that total adherence to the integrated supply system concept can divert the lion's share of the effort toward the decidedly less valuable portion of the inventory investment. There are, of course, long range, offsetting advantages in favor of segregating inventory management on a commodity basis. The relative value of these advantages at the outset of weapon system support of vital programs being worked against compressed time schedules are deserving of continuing consideration.

The U. S. Army system of support for weapon systems more closely parallels that of the Navy than does the Air Force. The Army Missile Command, consisting of the Ballistic Missile Agency and the Rocket and Guided Missile Agency has what is comparable to the technical bureau and program support supply demand control point responsibilities. Supply support of common items is negotiated with the several Army Technical Services (Signal, Engineer, and the like).

THE NEED FOR WEAPON SYSTEM MAINTENANCE POLICY

The beginnings of material logistic support requirements of weapons systems come rather early in the development of weapons systems. Since no one has ever guaranteed or could guarantee that the equipments of a weapons system will be failure-proof throughout, then the design criteria for the equipments must include some provisions for their maintenance. The level of maintenance may be a function of the availability of technical skills in various operating echelons. The levels of maintenance may be a function of capacities of test equipment and other physical facilities to isolate the failure and to give an indication of successful repair. These things are indications of the possible range of repair parts which may be required to maintain the equipments of Naval Weapon Systems on combatant ships, tenders and repair ships, in Naval industrial activities and at manufacturer's plants.

It is necessary, therefore, in the design of the several equipments which constitute a weapon system to give consideration to the maintainability of the equipment. When the consideration of maintenance is careful and complete, the complete, clear maintenance policy is easily evolved and the selection of the range of repair parts is greatly facilitated.

PROVISIONING DOCUMENTATION AND STANDARD PARTS

There are several objectives in the preparation of weapon systems which, when pursued independently are at cross purposes with each other and inventory policies which have been established in the Navy Department and in the Department of Defense. These are:

1. The enthusiastic proponents of complete provisioning documentation have tended in some services to deluge the equipment manufacturers with such ponderous repair parts paperwork that the manufacturers are rightfully asking the service customers to decide which they want more—paperwork or hardware. Obviously, in selecting repair parts for new equipment it is always desirable to have a few more facts before making a repair-part decision. Just as obviously, there is a point beyond which the pursuit of additional facts (or documentation) is not justified for small incremental increase in decision effectiveness which is attained.

Great strides have been made in recent years toward providing the necessary documents for the orderly, timely, and complete provisioning of equipment repair parts. More, however, can and must be done to keep repair parts provisioning specifications in step with the rapidly changing technology of weapon system. Military specification P-21078A is one example of a provisioning documentation effort for a maximum of useful provisioning information at acceptable cost.

2. A second perplexing problem lies in the fact that equipment designers and producers are under great pressure in these times to deliver workable products on a compressed time schedule so as to meet early installation dates in new construction and major conversion ships. The equipment designers are required by specifications to use "standard" parts wherever possible. The advantages of standardization are well known for the benefits to inventory managers through the reductions in inventory range. The designers' cry is that they are compelled to design around existing components. Their position is that they cannot design tomorrow's weapon systems with yesterday's components. Obviously, these diverging view points can be reconciled only by compromise.

PROGRAM SUPPORT BY A NAVY INVENTORY MANAGER

The logistic support of Naval Weapons System makes mandatory a strong and effective system of program support and supply support supply demand control points. Program support of all the principle equipments of a weapon system in a single SDCP is the only sound approach to the support of weapons systems from an integrated Navy Supply System.

It must be recognized at the outset that weapon system managers are not generally supply oriented and Navy Supply Systems conscious. This is not because the Navy Supply System is a thing of mystery and comprehensible only to people who draw their pay from an activity under the management control of BUSANDA. It is because the weapon system

managers devote their attention to the development and production of weapon systems which meet the operational requirements of CNO. This does and should require the full attention of the weapon systems manager. He generally has neither the time nor inclination to become intimately familiar with the mechanics of the integrated Navy Supply System. It matters little to him that there are such things as cognizance symbols, federal catalog requirements, and mandates against dual stocking of material in the Navy Supply System. He is, in a word, impatient with this relative trivia.

This condition is either unalterable or so nearly so that to do other than live with it would require such a prodigious supply orientation program that it is unlikely to be accomplished. Under the circumstances the wiser course is to concentrate upon developing a general confidence among weapon system managers for the Navy Supply System. This confidence is most easily realized through dealings with one responsive program support inventory manager of the Navy Supply System for each weapon system.

It is then the special responsibility of this office (in this case, the program support SDCP) to receive, utilize and disseminate all the necessary logistic, maintenance, and technical information for all the equipments of the weapon system to the rest of the supply system.

How the program support supply demand control point accomplishes these difficult tasks is a matter in which the supply system will be forever trying to improve itself. Safeguarding against dual stocking by good material identification, cataloging and supply support, maximum utilization of assets, and financing the inventory are all vexing problems in that their solutions take time. Time is a commodity that is always in short supply in trying to assure adequate range depth and quality of repair parts for a weapon system against a compressed schedule for an operational capability.

It is to be recognized that, by the time the customer begins requisitioning repair parts for a weapon system the preponderance of requirement-determination decisions have already been made. In order to insure the availability of repair parts by the time the equipment is operational, initial delivery of on board repair parts and supply-systems stock must be concurrent with the equipment delivery. When this requirement is coupled with the requirement to buy repair parts for which there is no other requirement than the weapon system, it is obvious that repair parts must be ordered when the equipments go to production. At that time, there is no demand history from which to project range and depth requirements. Yet, it is at this time that all the requirements decisions must be made. There is no opportunity to test the validity of these decisions until the equipment goes into operation.

WEAPON SYSTEM PROVISIONING

Historically the initial range and depth of equipment repair parts has been established by a spares technician who tries to recall experience with similar parts and who supplements his experience with intuitive judgement as to what items and how many of each should be procured. He has been required to consider such factors as consumption rate and variability,

cost, weight and cube, and compensability and shortage cost. In truth, however, many repair part technicians do not even comprehend the meaning of these terms.

Efforts are being made to set down an orderly pattern for repair parts decision making. The first step is a mathematical model which gives balanced attention to all the factors which influence depth of buy decisions in provisioning. This work is being conducted under Contract Nonr 2713(00). The ambitious goal of this program is a balanced load of repair parts which will fit within the available stowage cube of the Naval ship. The decisions as to what and how much material is to be carried in a combatant ship cannot be made logically without consideration of the type and frequency of resupply of deployed vessels. Obviously the economics of inventory management must be considered at this point.

One fact remains unalterable in the treatment of any weapon system. That fact is that the task of requirements determination of repair parts for new equipments is a most difficult one with which to cope. A sound approach to new provisioning techniques is in the making with the development of parameters and the determination of parameter values. Some of these parameter values are known now or can be assigned for weapon systems and equipments within weapon systems. Unit price, reorder cost, holding costs are relatively straight forward.

Penalty cost or the cost of equipment downtime is more nebulous. The absolute value of a sustained operational capability of a given weapon system in the scheme of national defense is not easily determined. Weapon system managers have several methods at their disposal to determine the probable inherent reliability of a weapon system. Whatever this inherent reliability may be, it must be assumed that it may be brought up to a figure approaching 100-percent reliability by the provision of the proper depth in the necessary range of repair parts. In substance, the proper maximum amount to invest in repair parts is the difference between the value of the offensive or deterrent capability to our defense position and the raw costs of development, equipment production, and manpower and facilities required for the system. Obviously, the investment in repair parts approaches infinity as the reliability approaches 100 percent. While 100-percent reliability and on-line availability is desirable as a goal, any sober consideration of the situation leads to the conclusion that something less than 100 percent is all that is attainable within the funds which may possibly be made available. Thus, it appears that military strategists must make the initial determination of the total value of each weapon system. After the best efforts have been exerted by the design and production people to deliver the system which has the best attainable inherent reliability, the difference remaining between that total cost and the total value of the weapon system should be the maximum amount available to operate and maintain the system. It is realistic in this manner to assign a fixed maximum dollar figure for the repair parts investment.

It remains, however, that an accurate determination of projected requirements must be made. The most essential ingredient in this determination is a forecast of the useful life of each repairable and replaceable part in the equipment. As the complexity of equipments increases, it becomes increasingly important to design equipments for easy maintenance by

the service personnel. In order to hold equipment downtime to a minimum, self-checking features must be incorporated in the equipment which will isolate the failure to a specific part or component. Then this component must be designed for universal interchangeability with the replacement part or component. The trend in weapon-system equipments design is toward the replacement of subassemblies instead of piece parts.

This trend complicates the determination of the required depth of replaceable parts. As the size and complexity of parts increases, the cost increases. In the past overestimating requirements for individual electronic piece parts, for example, was not relatively significant because of the comparatively low cost of parts and the high probability of the excess being consumed in other equipment applications. In today's context, the repair part is a group of 50 to 100 or more electronic piece parts assembled in a plug-in printed circuit module which probably costs between several hundred and several thousand dollars.

Relatively little has been done to develop techniques for estimating probable failure rates for modules which are in most cases single equipment application items. To overestimate parts requirements in today's equipments is to generate large stocks of high value material which must be scrapped when the equipment is superceded. The need is indicated, therefore, to do an accurate job of estimating requirements by determining the probable frequency of component failures in the terms of the periods which the equipment may be operating. Failure rates stated in terms of hours of operation fit the requirements of equipments which operates regularly or for extended periods. Fire control, launchers and test equipment fall into this category. Failure rates for components of the missile itself are best expressed as the expected failures for a given number of tests whenever the missile operates only periodically for test and check-out prior to firing.

In the case of both flight hardware (missile) and capital equipment failure data, these data must be estimated until there are empirical data generated to replace them. These estimated rates must be developed from the known functional behavior of the bits and pieces of which the modules are manufactured.

Printed-circuit electronic modules, for example, may be made up of 50 to 100 or more resistors, capacitors, diodes, transistors, and coils in various combinations. It is necessary, at the outset of the equipment production to develop arithmetic sums of weighted failure rates of the individual pieces to estimate the probable failure rate for the whole module.

This technique is obviously time consuming and the cost of this approach is prohibitive for the lower cost modules. Its value exists for higher cost modules, particularly where the state of the art indicates that there is a good spread in the useful life of the several components.

The value of this approach was demonstrated dramatically for a small number of very high cost components in the initial provisioning of Talos missile plug-in units. The contractor's first estimate of repair parts cost for USS GALVESTON (CLG-3), was approximately $4 million. Failure rates were subsequently developed for the three highest cost items. The

reduced requirements for the three items reduced the total cost of all Talos components down to $0.9 million.

The rank and file of equipment specialists who make the initial requirements determination are inclined to reject this technique in favor of their own intuitive judgment. The need is indicated for an education program which leads to every-day use of this tool for formalizing the thought processes which go into first, the selection of the item and, second, the procurement depth of the item.

The need is also indicated for developing techniques for distinguishing between consumable and repairable repair parts. This need is also an outgrowth of the trend toward plug-in units of repair parts for new weapons. The trend toward modular construction is accompanied by the trend toward miniaturization of piece parts and modules. There is a logical inclination toward the consideration of modules as repairable because the failure is of one or more of the piece parts in the module. The inclination is that if the faulty part or parts are replaced, the module is repaired.

In many instances this is true. However, it is also true that the cost in time and effort to isolate the fault and correct it oftentimes approaches the replacement cost of the whole module. There are investigations underway to develope trade-off problem criteria for distinguishing consumable items from repairable items. One such study is being conducted by TEMPO for the Special Projects Office.

The criteria for this type-decision-rule contain many variables. One of the more perplexing of these is occasioned by the trend toward miniaturization, particularly in electronic equipments. The present state of the art includes a capability to compress 100 of today's electronic bits and pieces into a space of 1-cubic inch. It is quite likely that in 10 years miniaturization shall have been advanced to the point that 100,000 of today's pieces will be compressed into a space of 1-cubic inch.

When the art has advanced that far, one thing will be quite clear. That is that there will be no point in designating the miniature module as repairable, since the whole thing will be less large than a lady's watch. Until that time, there is a need for "fix or replace" decision rules which consider space, both work and storage, training for the crew, and piece parts inventory on the one hand and the space to stow replacement modules and their higher cost on the other.

We have barely scratched the surface in this all-important area of provisioning. The Navy Supply System, including the technical Bureaus as principal item inventory managers, need, very much, a set of decision rules which will assure a consistently correct approach in this area which draws down so many material dollars.

MAINTAINING LOGISTIC SUPPORT

Up to this point, the discussion in this paper has been directed toward first-stage logistic support, that of the logistic planning and determining material requirements. It is clear that much of this work is accomplished on the assumptions that a certain set of operating conditions will obtain, that operating forces will utilize the weapons system furnished

them in accordance with instructions furnished them, and that the weapons system will function in accordance with engineering estimates.

At the present time, most of the "fact" input to tomorrow's weapons is based upon the projections of trends established in yesterday's weapons. It is too much to assume that the logistic package will be entirely suitable and that the weapon systems will operate without further guidance and tailoring where necessary. The need is well recognized for a feedback of information on operating performance of the equipment itself and the logistic support system.

The first need for such information is to correct deficiencies in the weapon system.

Some of the deficiencies are of such magnitude that they will require engineering changes to the equipment. More of them can be eliminated by adjustment in logistic support (e.g., adjustments in range and depth of repair parts) and increased tempo of resupply.

The normal mechanics of the replenishment requisitioning process, given enough time, will restore the balance in the logistic support equation which may be in imbalance as the result of unpredictable factors in the equation. When the consumption rate of repair parts is higher than that which was anticipated, more frequent ordering of the repair parts temporarily relieves the situation. However, without a feedback of information on the operating performance of the equipment involved (failure reports) the imbalance is not fully satisfied. Abnormal consumption of repair parts presents a two pronged problem.

First, abnormal consumption of repair parts is an effect and does not, in many cases, pinpoint the cause. The flat tire and the blown fuse are effects which presuppose an earlier cause of tube failure and an overloaded circuit, respectively. Similarly abnormal mechanical failure in weapon equipment presupposes faulty lubrication, strain, or some other cause; and abnormal electronic part failure may be caused by a fault elsewhere in the circuitry. Without investigating the causes of equipment failure thoroughly, the continuous replacement of failed parts is as unproductive as bailing out the boat without repairing the leak. Failure reporting is the first step in the cycle which is not complete until the reporting is followed by investigation, analysis and correction. However, failure reporting is the very important first step. In the Naval Service, the operating failures occur many miles at sea, a long distance from the laboratories and factories from which the equipment is produced and where the investigations and analysis of operating failures will lead to satisfactory correction of the faults.

If the failure reporting system fails, and many have, it is usually because the work falls to operating personnel who are already overburdened with paper work and who usually see no tangible result of their reporting effort particularly "on their watch." It follows, therefore, that failure reporting systems must be simple. Reporting formats must be concise enough to be completed quickly without being meaningless. The failure-reporting program should continue only long enough to serve its purpose. It should be limited in scope to the reporting of significant failures. The efforts of failure reporters should be rewarded with a timely feedback of failure analysis and corrective instructions.

The second problem is related to the first. The Navy logistic system is based on onboard allowances of repair parts, replenishment stocks in tenders and supply ships, back up stocks in the several echelons of the supply system ashore and a procurement program to replenish the pipeline. Abnormal consumption on board ship causes abnormal demands through the whole system. The first unsatisfactory condition is a rash of not-in-stock positions in the logistic pipeline at all levels. More frequent ordering of the parts in question provides partial relief. However, when the onboard allowance and tender/supply ship stock levels are not adjusted, the second unsatisfactory condition obtains, that of inadequate endurance. The need, therefore, is obvious for a feedback of usage information to effect an upward adjustment of stock levels afloat to compensate for the increased consumption.

In the case where there is a feedback of failure data and usage data to the equipment engineers and supply program managers, respectively and independently, a more complicated interaction may develop. At the time the equipment engineers are working toward more reliable components to reduce equipment downtime, the supply system is usually procuring more of the failing parts in response to demand data.

Again, it is obvious that there must not only be a failure-reporting system and a usage reporting data system, but the two systems must be coordinated. The supply program manager must know what the equipment engineer is planning and doing in order to avoid stocking the now obsolete part and to stock, as necessary, quantities at all levels of the superceding part. Where the correction of equipment deficiencies is formalized in an alteration program the supply manager must participate to the extent of deleting and disposing of superceded parts and providing the replacement part. There is no assurance of responsive support from the supply system if the alteration program is not formalized to the extent that the supply program manager participates in it.

At this point it is desirable to review the process the supply system follows placing the initial repair-part support for equipment in ships and the means of sustaining support for equipments. The COSAL program includes an allowance parts list section (APL) in each COSAL. APL's for new equipments are usually developed through the provisioning process. Each ship is now, or will be in the near future, provided with a new COSAL incident to construction or major conversion. This COSAL will be revised at the time of periodic ship overhaul.

The APL's for new equipments are, as has been discussed, the result of a series of best estimates of part failures and are the best estimates of what will be required to support the installed equipments. If the COSALS are to be kept in step with the actual requirements of the equipments they support, the APL's must be revised to reflect the repair parts usage experience which comes from a usage data program. This writer's knowledge of all inventory managers' usage data and APL maintenance programs is admittedly incomplete. However, 40 months of more than casual inquiry in this area leads to one person's conclusion that while interested and alert to the problems, the inventory managers usually do not have the time or funds to do a comprehensive job of updating the range and depth of APLs for COSALS.

Excellent progress is being made in the area of supply availabilities at the time of ship overhaul. Armed with a new COSAL for the ship, the supply-availability team purges out repair parts which were applicable to equipment which has been removed. Stock records, bin tags, and locators are brought up to date to reflect stock number and cognizance changes. The big questions are several. Has the quality of the APL's in that COSAL been improved in range and depth to reflect the historical repair part demand data? Have quantities been increased where demands have been higher than the original COSAL estimate? Have quantities been reduced where equipment part reliability is higher than originally anticipated? Are new items being added on the basis of demands which could not have been anticipated in the original COSAL? If the answers to these questions and a battery of others like them are negative, the overhauled ship goes back to sea with a repair part allowance which has been brightened up with new labels but without substantial qualitative improvement.

The first emphasis on ship's allowances has been to ensure that all the repair parts called for in the allowance lists are on board. This has been done without substantial challenge to the validity of the lists themselves. Unquestionably, ships must carry necessary spares. There are also two important reasons why ships should carry only those repair parts they will need and can use. The first is money. In a climate of fixed budgets, investments in excessive range and depth of unnecessary parts prevents the necessary investment in needed parts or limits the number of ships which can be so outfitted. The second reason is a space constraint. As ships and their weapon systems become more complex the number of required repair parts increases. As these new weapon systems are placed on more smaller ships (e.g., destroyer and submarine types), the task of finding places for these repair parts becomes more complex. The USS TRITON reportedly has repair parts stored in 900 locations in the ship.

Weapon system reliability, dollar constraints, and space constraints are irrefutable arguments in favor of continued surveillance and modernization of Allowance Parts Lists.

If we can accept as a fact that it is necessary to update APL's systematically based upon historical usage data, the problem is how and where to begin when it is well known that the inventory managers are already hard pressed to keep up with the job of writing new APL's for new equipment. It is manifestly impractical to expect that it is possible to update all existing APL's. The key to approaching the task lies in concentrating first on the high military worth equipments.

Military essentiality studies have been instituted as a means of ensuring first that critical repair parts for critical equipments are always available. Military essentiality is subdivided into mission effect and compensability. Mission effect designations are the means of ranking the importance of equipments to the accomplishment of the ship's mission. It is submitted that, with certain modifications, it is entirely practical to use these mission effect designations as a means of determining the order in which weapon system APL's should be updated.

NAVAL WEAPON SYSTEM LOGISTIC AGENCIES

There is discussion earlier in this article of the general pattern of material logistic support of weapon systems. The most recent major departure from the general pattern of Naval weapon system support is in the Fleet Ballistic Missile System. The Chief of Naval Operations directed in March 1959 that CINCLANTFLT establish a Fleet Polaris Material Office. This office is charged with responsibilities for processing FBM requisitions, exercising issue control over critical FBM Material, recommending changes in allowance lists and maintaining liaison with supply activities and inventory managers.

It should be noted that the PMO is a Fleet activity. Its job is to ensure that the fleet's requirements for the sustained support of the FBM weapons system are relayed to, understood by and satisfied by the shore establishment. The PMO is not charged with the functions of supporting the FBM weapons system. These are the functions of the Shore Establishment. In a sense, the PMO is the Fleet voice in announcing the what, when and where of required support.

The value of the PMO rests in the fact that its attention is directed entirely toward the FBM weapon system and toward the entire FBM weapon system. It is intended that the PMO will be expert in knowing all FBM requirements. One of its most effective tools for doing its job is a thorough understanding of the support contributions of all the inventory managers. It will receive the fleet support requirements from the operating FBM units and pass them in supply language to the highest levels necessary in each inventory manager's distributive system to satisfy these requirements.

As the logistic agent ashore for the FBM fleet, the PMO must ensure that the logistic requirements of that fleet are understood by the shore establishment and satisfied as stated. It should not be inferred from the foregoing stated objectives of PMO that its establishment is a denial of the integrated Navy supply system. The establishment of PMO is recognition of several significant facts: first, that within weeks of the delivery of operational missiles the first SSBN will be deployed on operational patrol with its strategic target assignment; second, that once operational, each fleet unit will be deployed outside CONUS continuously until overhaul; and third, that the strategic value of the FBM weapon system varies directly with the capability to keep the weapon system "up" or operational at all times.

The FBM weapon system is a most ambitious undertaking with the promise of being one of the most effective deterrent weapons in the United States military arsenal. The PMO is one of the devices which will contribute to the logistic assurance of this capability. The value and urgency of the FBM weapon system to this country's military posture justifies the establishment of the PMO.

In accepting the concept of the PMO for the FBM weapon system and assuming the substantial contribution which is expected of it in the support of FBM ships, shouldn't the concept be extended to other Naval weapon systems? It is submitted that the answer is probably negative. The FBM program presents a pattern of considerations which are singularly adapted to the PMO concept (i.e., identical capabilities in all ships, single and

identical assignments, and similarity of installed equipments). New construction surface ships are being configured with several combinations of surface-to-air, surface-to-surface and surface-to-underwater weapon systems. Terrier, Talos, Tartar, ASROC and others are separate weapon systems which will be arranged in various combinations in CG, DDG, DLG, and CV type ships. The individual differences in ships themselves, the schedule for their completion and fleet assignments present a condition which is not conducive to improving fleet support by a PMO-type logistic agent.

Surface ships which are commingled in various combinations for various periods in striking forces are probably susceptible of most effective operational support through the existing logistic channels.

However, many of the support problems in initial outfitting and reoutfitting are present in surface ships which are being met for SSBNs through the PMO. The Ordnance Supply Office has recognized—for purposes of initial outfitting—that each ship, with its combination of armament, is a weapon system. OSO has been, for over a year, assigning hull coordinators (for such types as DDG, DLG, and CG) to see each ship through initial outfitting. The Aviation Supply Office, similarly, has provided coordination of initial support of new aircraft types.

Many of the elements of weapon system support such as integrating individual provisionings of equipments which comprise a weapon system, COSAL development, material distribution and stock positioning, and filling initial allowances require a certain amount of coordination as a weapon system in order to be handled effectively. Usage data collection and analysis for the purpose of maintaining APL's for COSALS cannot be effective except on a weapon system basis.

It is submitted that the program support Supply Demand Control Points are best suited to this type of work and can accomplish it with the least amount of modification of organizations and make ready time. What would be required is a positive communication channel to and from the fleets to bring the inventory managers closer to the operators on a weapon-system basis. An expansion of the hull coordinator and aircraft support liaison concepts of OSO and ASO, respectively, appears to be in order. It does not necessarily follow that it is necessary to reorganize personnel and add new organizational units for this purpose. It would seem more desirable to program as many as possible of the functions of sustained weapon system support on electronic data processing equipment. For example, mathematical decision rules are now being applied to weapon system provisioning. A by-product of this exercise is the equipment APL for the fleet COSALS. It follows that usage data and failure data (actual) could be fed into the computer to redetermine the previously estimated onboard range and depth of repair parts.

A program of sustaining the support of equipments on a weapon system basis in the program SDCP's would effectively "close the loop" to the operating forces for weapon systems. It would provide the means for feeding back the data to evaluate and improve initial provisioning decisions. It would be the means of insuring in the supply availability program

that ships would receive true onboard allowance overhauls as well as equipment overhauls. A demonstrated capability in the SDCP's to perform these functions would do more than anything else to preclude proposals for the establishment of special logistic offices to support new programs and weapon systems.

SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS

Actual experience in several branches of the armed forces has demonstrated the need for concentrating the program support of each defined weapon system in a single organization. It is recommended that the Navy Supply System continue to focus upon the needs of the operating forces and the specific requirements of weapons system in making weapon system program support assignments to inventory managers.

Clear, concise statements of weapon system operation and maintenance, when stated in a maintenance policy are the best guide lines for program support SDCP in picking up logistic support. It is concluded that whenever one is already not provided, the Inventory Manager should press the responsible technical bureau for maintenance objectives stated as a policy by the time of the first system equipment provisioning.

There are a number of justifications to the fleets, technical bureaus, SDCP's and equipment contractors for the early adoption of computer processing of mathematical provisioning rules to raise, uniformly, the quality of provisioning decisions. It is concluded that BUSANDA should continue to press for the acceptance of mathematical provisioning decision rules by technical bureaus and program supply demand control points. Pockets of resistance to these rules are present. The resistance is usually born of unwillingness to try to understand the rules. An education program is necessary—but, as time goes and as the program gains favor, the remaining resistance should be treated with calculated impatience.

Directly related to the provisioning decision rule conclusion is another concerning the assignment of parameter values for making decision rules. Not only are parameter value assignments necessary for completing the mathematical calculation, they are invaluable as a technique for demanding that the equipment designer, builder, and provisioner alike sharpen their thinking on repair part decisions and thereby do a much better job. It is recommended that the technical SDCP's and bureaus who have provisioning responsibility give early and sustained emphasis for defining parameters and setting parameter values for provisioning.

The value of failure and usage reporting is well recognized. On the other hand the inventory managers and others responsible for maintaining allowance lists have very little hope of maintaining, through usage data all APL's.

It is recommended that the mission effect element of military essentiality of equipments be used as the criterion for preference ranking of equipments for the scheduled maintenance of APL's on these equipments.

The establishment of the Polaris Material Office is justified on the basis of the unusual urgency and importance of the Fleet Ballistic Missile Program. It does not

necessarily set a precedent for all subsequent weapon systems. It does fill what could be a void in subsequent weapon systems unless an alternative is adopted.

It is recommended that the program support supply demand control points become the material offices for new weapon systems by providing the means for "closing the communications loop" with the Fleet on a weapon system basis in order to insure that adequate support of weapon systems is provided through the life of these systems.

BIBLIOGRAPHY

[1] Report of the Committee on Organization of the Navy Department.

[2] "Inventory Management Review of the Navy Special Weapons and Guided Missile Program," April 1959, Office of Naval Material.

[3] Commander K. M. Beyer, SC, USN, "Support of the Fleet Ballistic Missile Program," March 1959.

[4] "Undersea Support and the Navy's Future," Armed Forces Management, April 1959.

[5] "The Role of Maintenance Analysis, Polaris Fire Control System," RM 59TMP-8, January 30, 1959, Technical Military Planning Operation, General Electric Company.

[6] "Dependability of Military Equipment: A System Approach," Electrical Manufacturing, March 1959.

[7] "Report on Phase A Polaris Fire Control System Logistic Management Study," RM 59TMP-8, January 30, 1959, General Electric Company.

[8] Recommended Repair Parts List, Talos Type II Missile.

[9] Recommended Repair Parts List FBM Shore Support Equipment, NLR 30000.

[10] OPNAV NOTICE 5440, 17 December 1959.

[11] Provisioning Requirements for Support of Naval Ordnance Equipment MIL-P-21078.

* * *

A NOTE ON PYRAMID BUILDING*†

Zivia S. Wurtele

It is of interest for planning purposes to determine the effect on the rate of balanced growth of adding to an economy industries whose outputs are used neither directly nor indirectly for the production of consumption goods. Examples of such industries may be found among defense industries; and a classic example is Keynes' pyramid building.

Let us assume that we have a closed Leontief-type economy which is divided into n sectors and which is lagged so that the outputs of one period are the inputs of the next. Let A and B, respectively, be n-order matrices of technical coefficients a_{ij} and capital coefficients b_{ij}. Let x^t be a column vector of the outputs x_i^t (i=1, ... , n) of the sectors during the period t. We assume that (I-A) is a Leontief-type decomposable matrix and that the industries of the economy have been aggregated into sectors so as to achieve a matrix of technical coefficients of the form:

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ * & A_2 & & \\ & & \ddots & \\ * & & & A_k \end{pmatrix},$$

where A_e (e=1, ..., k) are square, indecomposable matrices of order n_e, with

$$n = \sum_{e=1}^{k} n_e,$$

where all entries above and to the right of A_e are zero, and where the last sector of A is labor. Thus the sectors in the e^{th} block A_e, for $e < k$, are needed neither directly nor indirectly for consumption. Similarly, let B_e be the submatrix of capital coefficients corresponding to the coefficients of A_e, e = 1, ..., k.

The equations of our system, lagged so that outputs of one period are inputs of the next, are:

$$x^t = A x^{t+1} + B\left(x^{t+1} - x^t\right) \tag{1}$$

Setting $x^t = (1+d)^t x^0$ in Eq. (1), we obtain:

---

*Manuscript received July 1, 1960.

†This work was supported by the Office of Naval Research, Contract Nonr 233-02, under Task NR 047-003, Management Sciences Research Project, University of California, Los Angeles. Reproduction in whole or in part is permitted for any purpose of the United States Government.

$$\left(I/d - (I-A)^{-1}(A+B)\right)x^0 = 0. \tag{2}$$

By means of some fairly straightforward algebraic manipulation, it can be shown that Eq. (2) is of the form:

$$\begin{bmatrix} \left(I/d-C_1\right) & 0 & \cdots & 0 \\ * & \left(I/d-C_2\right) & & 0 \\ & & \ddots & \vdots \\ * & & & \left(I/d-C_k\right) \end{bmatrix} \begin{bmatrix} y_1^0 \\ y_2^0 \\ \vdots \\ y_k^0 \end{bmatrix} = 0, \tag{3}$$

where the matrices $C_e = (I-A_e)^{-1}(A_e + B_e)$ are indecomposable and y_e^0 (e=1, ..., k) is a vector of the outputs $x^0_{n_1+n_2+\cdots n_{e-1}+1}, \ldots, x^0_{n_1+n_2+\cdots n_e}$ of the sectors of the e^{th} block.

Let $1/d_e$ be the maximal characteristic root of C_e; and assume that if $m \neq e$, $d_m \neq d_e$. We now make the additional assumption that the matrix of coefficients in Eq. (3) is not separable. (The separable case may be studied by first dividing the matrix into nonseparable submatrices.) The elements of C_e are positive. If Eq. (3) holds for nonnegative y^0, then d must equal d_e for some e such that $1 \leq e \leq k$.[1] Suppose $d = d_j$. Since $\left| I/d_j - C_1 \right| \neq 0$, we must have $y_1^0 = 0$. Since $y_1^0 = 0$ and $\left| I/d_j - C_2 \right| \neq 0$, we must have $y_2^0 = 0$. In this way, we may prove that $y_h^0 = 0$ for $h < j$. Since C_j is indecomposable, $y_j^0 > 0$. Consider now the e^{th} set of equations of Eq. (3) where $e > j$. These can be written in the form:

$$y_e^0 = \left(I/d_j - C_e\right)^{-1} z_e, \tag{4}$$

where z_e is a nonnegative vector. A necessary and sufficient condition for the solution y_e^0 to be positive is that $d_e > d_j$.

Suppose that to an economy consisting of the sectors of the j^{th} through k^{th} groups, with balanced growth rate d_j, are added the sectors of the $(j-1)^{st}$ group. There are two cases to distinguish. If $d_{j-1} \leq d_j$, uniform growth is possible with positive outputs for all sectors of the enlarged economy, and the new growth rate is d_{j-1}. On the other hand, if $d_{j-1} > d_j$, uniform expansion with positive outputs for all the sectors in the $(j-1)^{st}$ through k^{th} groups would be physically impossible. In evaluating a plan for enlarging the economy by the addition of industries whose outputs are not used directly or indirectly for consumption, it is thus appropriate to determine whether it is compatible with balanced growth and positive outputs for all the sectors of the economy.

---

[1] The argument of this paragraph rests heavily upon the results in [1] and [4] .

BIBLIOGRAPHY

[1] G. Debreu, and I. H. Herstein, "Nonnegative Square Matrices," Econometrica, Vol. 21, No. 4, October, 1953, pp. 597-607.

[2] J. M. Keynes, The General Theory of Employment, Interest, and Money, 1936.

[3] W. W. Leontief, and others, Studies in the Structure of the American Economy, 1953, 561 pp.

[4] M. Woodbury, "Characteristic Roots of Input-Output Matrices," Economic Activity Analysis, O. Morgenstern, ed., 1954, pp. 365-382.

[5] Z. S. Wurtele, "Equilibrium in a Uniformly Expanding Closed Leontief-Type System," Review of Economic Studies, Vol. 28, No. 1, October 1960.

* * *

AIRCRAFT COMPARTMENT DESIGN CRITERIA FOR THE ARMY DEPLOYMENT MISSION*†

W. F. Sharpe

University of Washington

An important mission of military cargo aircraft is the deployment of army equipment to areas of actual or potential warfare. The large number of vehicles required to support army forces makes weight and cube, the traditional measures of airlift capacity, insufficient for analysis of the amount of such cargo which can be carried by various aircraft. This paper uses a representative list of army deployment cargo to investigate the capacity of alternative aircraft designs. A loading model is developed to estimate the relationship between cargo compartment dimensions and loading efficiency. The relationship between compartment floor area and floor area of vehicles loaded is found to be remarkably stable over a large range of alternative compartment sizes. A second model is used to minimize the distribution of total weight per aircraft when the aircraft are loaded in accordance with the criteria of the first model. The resultant distributions, presented for various compartment sizes, suggest the relationship between cargo compartment and payload capacity which would be optimum for this particular mission.

INTRODUCTION

Aircraft loading problems are intimately connected with design criteria. Transport design affects both the preferred loading technique and the maximum amount of materiel the aircraft can transport over any given route. One of the most persistent problems in air transportation has been how to determine the capability of cargo aircraft to perform various tasks. It has long been recognized that it often may be either impossible or impractical to load aircraft with the maximum payload allowable over the chosen route; when this is so, some measure other than weight must be used in calculating airlift capability. Cargo cubage appears to be an acceptable measure for the type of materiel airlifted by commercial carriers and for most equipment moved by the military in peacetime; in both cases, most of the cargo consists of small units which lend themselves to flexible arrangement, so that the cargo compartment space can be efficiently used in all three dimensions. Under these circumstances, cubage may be used as the sole measure of space requirements in airlift planning.

The important relationship for peacetime cargo is between its weight and its cube. If an aircraft's allowable payload over a route exceeds the weight of the cargo which will fit into the cargo compartment, cubage is the constraint on airlift capability.

*Manuscript received October 3, 1960.

†This study was performed while the author was a member of the Logistics Department of the Rand Corporation.

If the allowable payload is less than the plane could carry, weight is the constraint. Considerable research has been devoted to this important problem area and the implications for airlift planning and aircraft design have been examined in some detail.[1]

Unfortunately, the procedures for studying peacetime cargo airlift do not appear applicable to army deployment, which involves a different kind of cargo. The bulk of the cargo consists of vehicles which cannot easily be arranged to fill the cargo compartment. The cargo cannot be stacked to take full advantage of compartment height; most of them are longer than the compartment width, so they cannot be turned sideways; and it is impractical to turn them on their sides. Cubage alone, then, is not an adequate measure in this case. Detailed investigation of the space requirements of army deployment cargo is necessary in order to obtain a more accurate measure. Such a measure can then be compared with cargo weight and fruitful design criteria for this mission derived.

This article describes a method which can be used to attain these goals. To aid the exposition a particular application of the method will be described. We will examine the problems of loading a particular list of equipment: that required by a hypothetical planning force designed by the Office of the Deputy Chief of Staff for Logistics of the U.S. Army.* The conclusions reported here will apply to deployments similar to those of the force studied, of course. Moreover, although it cannot be said for certain that other deployments will have similar requirements, it is extremely likely that they will, since most of the vehicles studied are standard equipment for almost every sizeable army organization and make up a relatively constant percentage of any major force. The analysis should thus apply to a number of alternate mixes.

This article considers the following problem. There is a specified list of army deployment cargo to be airlifted. All the items can be placed in aircraft with cargo compartments at least 110 inches high, 120 inches wide, and 400 inches long. A number of aircraft designs are being considered for this task. Each has a cargo compartment at least 110 inches high but widths vary from 120 to 150 inches while compartment lengths vary from 400 to 1200 inches. Which design is best for this mission? What is the desirable relationship between compartment dimensions and payload capacity?

We will approach the answer to these questions in two stages. First, a loading model will be developed to estimate the relationships between cargo compartment dimensions and loading efficiency. Compartment width will be examined first, followed by compartment length and cubage. A second model will then be developed to minimize the distribution of total weight per aircraft when the aircraft are loaded in accordance with the criteria of the first model. The resultant distributions, presented for various compartment sizes, suggest the relationship between cargo compartment and payload capacity which would be optimum for the specified mission.

---

[1] See, for example, R. E. Bickner, Cargo Density and Air Transportation, The RAND Corporation, Research Memorandum RM-1380 (ASTIA Document No. AD 86398), May, 1955; and idem, Cargo Density and Airlift, The RAND Corporation, Research Memorandum RM-1853 (ASTIA Document No. AD 112417), January 14, 1957.

*A copy of this list of equipment was made available to the RAND Corporation in September 1959.

COMPARTMENT WIDTH

Compartment width claims our first attention; this section attempts to specify the most efficient method of using it in loading vehicles. The problems of compartment length are disregarded here, to be taken up in the following section.

For expositional simplicity, assume it is possible to build an aircraft of any length but with a specified width; given this constraint, we wish to determine the shortest possible aircraft which will just hold all the vehicles to be transported. The compartment is wide enough to hold some vehicles side by side, but no more than two, since the narrowest vehicle is more than one third the width of the compartment (this applies to all compartment widths studied, since the narrowest vehicle is over 50 inches wide).

The technique is simple analytically and could be applied on the flight line. The vehicles are first lined up with their left sides against the left wall of a very long cargo compartment; the widest vehicles go in first. Graphically, this is merely the cumulative column length as a function of vehicle width. In order to account for the space required for hold-down cables or straps, as well as problems of maneuvering the vehicles into aircraft, the vehicle dimensions have been increased. Both the vehicle widths and lengths reported here include this adjustment. Figure 1 shows the function for the vehicles in this study. The horizontal axis gives the column length which can be made up of vehicles whose widths exceed the corresponding value on the vertical axis. Column lengths are expressed in terms of the total column length (the length of the vehicle column when no doubling of vehicles is permitted).

Next, the last—and narrowest—vehicle pulls out of the line and drives forward, with its right side against the right compartment wall. The next-to-last vehicle follows it. The process continues until one of the vehicles can drive no farther because the space is too narrow

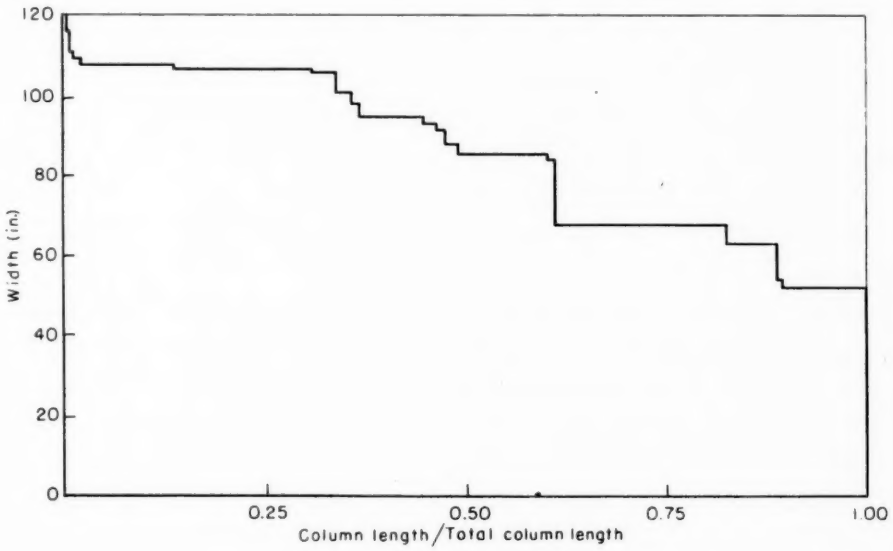


Figure 1 - Cumulative distribution of vehicle width

between the vehicle to its left and the compartment wall to its right. It is easy to perform this process graphically by using a transparent overlay of the original distribution: the overlay is rotated 180 , with the original horizontal axis placed on the horizontal line corresponding to the compartment width; the overlay is then moved from right to left until the two curves touch. Figure 2 shows the result of applying the technique to a compartment width of 150 inches. The total length of the new column is about 0.72 of that of the original. The minimum column length has been attained because the vehicle represented by the shaded area can move no farther forward (to the left in Figure 2).

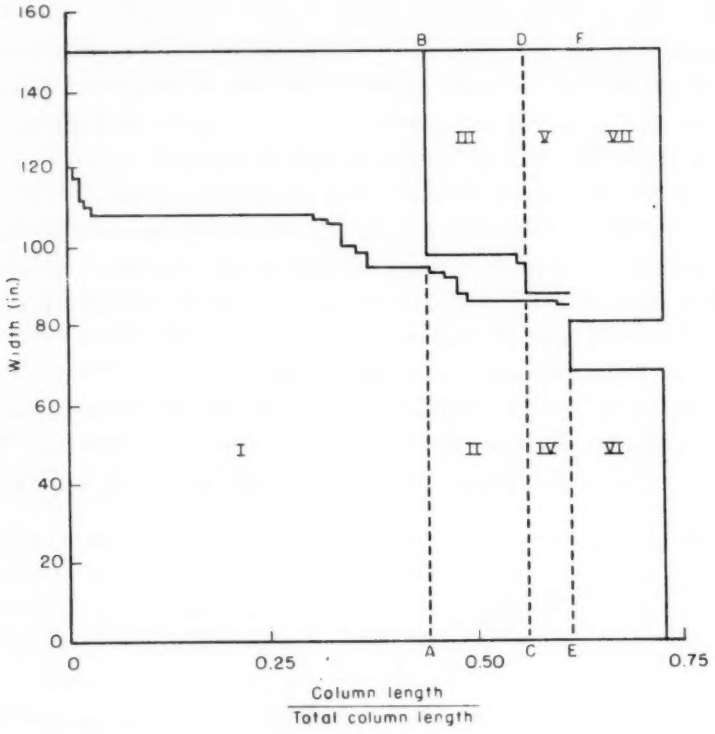


Figure 2 - Aircraft loading: compartment width 150 inches

An intuitive argument for the use of this technique is relatively simple. The vehicle represented by the shaded area cannot move to the left of the line EF since the original vehicles (Groups I, II, and IV) do not leave enough space. But the vehicles in Groups VI and VII are at least as wide as the vehicle in the shaded area since we have arranged them in order of decreasing width. Thus, they cannot be placed in the area to the left of EF. Therefore, the only vehicles which can possibly fit in this area are those in Groups III and V, which have already been placed there by the technique. Since no further rearrangement of vehicles can thus decrease the length of the column, the minimum column length has been attained.

By repeating this procedure for each of the compartment widths to be evaluated, we can find the corresponding minimum possible column length. This can be done in a matter of minutes once the transparent overlay of the initial distribution is made. In Figure 3 the solid line indicates the minimum column length associated with each of the aircraft compartment widths. The column length is again expressed as a proportion of the length of the initial column.

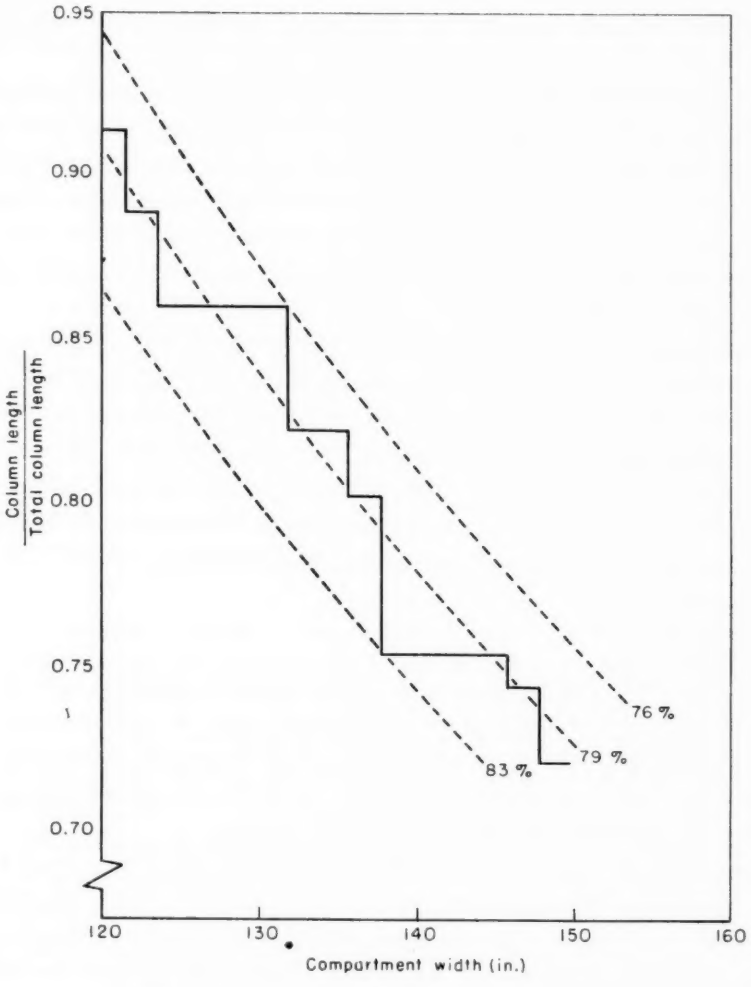


Figure 3 - Utilization of compartment width

The horizontal axis in Figure 3 represents compartment width, while the vertical axis represents the total length of cargo compartment(s) required for the entire group of vehicles. Any rectangular hyperbola of the form $xy = K$ represents all combinations of compartment width and column length which have the same total floor area. Three such functions are shown by the dotted curves in Figure 3. (Note that since the axes do not start at zero the hyperbolae appear almost linear.) We know the total floor area of the vehicles themselves, so any total compartment floor area can be related to the vehicle floor area by means of a load factor. Thus, a load factor of 76 percent means that the vehicles will cover 76 percent of the total floor area; stated in another way, the total floor area equals 1/0.76 times the required vehicle floor area.

For the vehicles included in this study, Figure 3 shows that the simple rectangular hyperbola provides an excellent equation for estimating the relationship between aircraft width and column length. The function everywhere lies between the curve representing a 76-percent load factor and that representing an 83-percent factor, and is well approximated by that representing a 79-percent factor. Within the range considered, we may conclude that aircraft width is irrelevant: 79 percent of the compartment floor area is covered with vehicles regardless of compartment width.

There is no a priori reason that a constant floor-area load-factor should adequately summarize the effect of varying compartment width. Its use will simplify the remainder of the analysis, but the analysis could also be performed with some less convenient estimating-equation for this relationship. Other lists of equipment might yield a curve which could be better approximated in some other manner. But the list under study here was selected because it was thought to be typical of army cargo to be deployed in contingency missions, and it is likely that many alternative lists would exhibit the same relationship between aircraft width and column length.

COMPARTMENT LENGTH

In the preceding section we minimized the length of the total column of vehicles to be airlifted, subject to various constraints on the compartment width. We concluded that if compartment length could be perfectly utilized, 79 percent of the total floor area could be covered with vehicles. This is an impossible ideal; to assess the real possibilities we must now examine the effect of compartment length on aircraft loading.

The same physical fact prevents perfect efficiency in using either compartment length or width: vehicles cannot be chopped into convenient sizes. But compartment length is more likely to be used efficiently than is width, because compartment lengths are greater in comparison with the average vehicle length than compartment widths are in comparison with the average vehicle width. Moreover, we would expect longer compartments by nature to be more efficient than shorter ones. Obviously, the approximate magnitudes of these relationships need to be determined.

There are no simple rules for the optimum use of compartment length; and in assessing the capability of various aircraft, no rule should be considered at all if it cannot be applied with some speed on the flight line. A further complication arises when we attempt to reconcile

a technique for obtaining optimum use of compartment length with the technique described in the preceding section for optimum use of compartment width. We shall initially disregard width (as we did length, in the preceding section) and return to it later.

Since there is no simple optimum rule for length, we will have to look for some simple and fairly efficient working rule. Tentatively, we will use the following technique.

We are given a group of vehicles to be loaded in single file, none side by side. The aircraft have a given compartment length. The goal is to load the vehicles but use as few aircraft as possible.

The vehicles are first lined up in order of decreasing length. The longest vehicle goes on the first aircraft, then the second longest, and so on until the remaining space is too short for the next vehicle in line. At this point, the longest vehicle which will fit is loaded.

The general rule is to load no more than one vehicle of any particular length on any single aircraft. The rule must be broken toward the end of the process, however, since one each of the few remaining vehicle lengths cannot fill the aircraft. In such a case it is permissible to load more than one vehicle of the shortest available length in a given aircraft.

When the vehicles are all loaded, we count the aircraft used; this quantity, times the compartment floor-length of the individual aircraft, gives the total compartment floor-length required. The total length of the vehicle column, divided by this total compartment floor-length, gives a loading-efficiency ratio with respect to compartment length.

The rule suggested has virtues other than simplicity. It imposes at least a small degree of combat loading; that is, it makes the loss of any one aircraft somewhat less crucial since the great majority of the aircraft contain only one vehicle of any given model. Further, it will tend to reduce the distribution of weight per aircraft.[2] Finally, it may prove fairly efficient.

To get a first approximation of the efficiency of this rule, it was tested on the full list of vehicles, assuming that none would be placed side by side. The longer compartment was more efficient, as we expected. Cargo compartments of 500 and 1000 inches were evaluated; the efficiency of the shorter compartment was 95.5 percent, of the longer, 98.1 percent. The size of these percentages indicates that the rule is surprisingly efficient and quite satisfactory for our purposes.

When we turn to a consideration of double columns of vehicles, we once more encounter the problem of compartment width, whose optimization was discussed in the preceding section. Doubling also compounds the length problem, however, since the end of a vehicle in one column may come in the middle of a vehicle in the other column. If the resulting loss of

[2] There are two constraints on aircraft capacity: the vehicle capacity (determined by the above analysis) and the payload capacity. The payload per aircraft loaded according to vehicle capacity is the ratio of these two dimensions for the equipment to be moved. The more this distribution clusters around its mean, the less inefficiency will arise because of the presence of the two constraints; and hence the smaller the number of aircraft of any given design which will be required to move the equipment.

length-efficiency is substantial, it may even be wise to consider less efficient methods of using compartment width; but as a first cut at this problem, we will see how efficiently length can be used if we retain the doubling techniques described in the preceding section.

The method adopted can best be seen with the aid of Figure 2, which shows the arrangement of vehicles in an aircraft 150 inches wide. In any aircraft, there will be some wider vehicles which will have no others alongside them (Group I, in this case). At the end of the column, there will also be a few pairs of vehicles with the same or nearly the same width (Groups VI and VII); we can regard these pairs as single vehicles and mix them at will with the wider vehicles (Group I). We can now apply our length-utilization rule to this composite group of vehicles to determine the number of aircraft required to hold them.

Next, we take the remaining vehicles and divide them into groups according to the following criterion: the vehicles can be rearranged within a group in any manner, without violating the width constraint. Groups II, III, IV, and V in Figure 2 are examples of such groups. Next, we apply the length-utilization rule to each such group, obtaining the number of aircraft required. The aircraft required for two groups which go alongside each other is merely the larger of the aircraft required for the individual groups. Thus, if Group II requires more aircraft than does Group III, the aircraft required for Group II will suffice to carry both groups. This technique allows us to determine a number of aircraft which can contain all the vehicles. We may express the efficiency of the technique by dividing the length of the initial (doubled) column by the total compartment length of the aircraft required. The results for the four extreme compartment sizes are as follows:

Compartment Length (in.)	Compartment Width	
	120 in. (%)	150 in. (%)
500	91.8	89.4
1000	95.4	91.2

Again the efficiency is surprisingly high. The narrower compartment uses length more efficiently since relatively few vehicles have been doubled, while the wider compartment pays for its efficient use of width by achieving a slightly less efficient use of length. The advantage of longer compartments is evident for each compartment width.

These differences are not insignificant; they suggest that a really sophisticated analysis might profit from a study which applied several possible techniques to this problem, evaluating the efficiency of such techniques for a large number of possible compartment sizes. The remainder of this study, however, adopts the very simple assumption that length can be used with 91.5-percent efficiency regardless of compartment size. This assumption has the virtue of simplicity and should not prove far wrong for most compartment dimensions under consideration, since compartment length and width are usually correlated. The efficiency of the

short, narrow compartment is 91.8 percent, while that of the long, wide compartment is 91.2 percent. If a compartment is to be considered which sharply diverges from these proportions, the final values obtained in the analysis should be adjusted appropriately.

We may now summarize the analysis of this section and the preceding section. In the preceding section, we found that vehicles would cover 79 percent of the floor area in a perfect utilization of compartment length; if the policies suggested in this section are adopted, however, only 91.5-percent efficiency appears obtainable in terms of length utilization. In that event, only about 72 percent of the floor area will actually be covered—91.5 percent of 79 percent. Stated another way, the minimum number of aircraft required to hold the vehicles can be computed by dividing the vehicle floor area by 0.72 to determine the aircraft floor area required. This figure is then divided by the floor area of the compartment size in question, to determine the required number of aircraft. And this technique applies to the full range of aircraft under consideration.

COMPARTMENT CUBE

Our analysis thus far has been concerned only with the vehicles to be airlifted. The study must now encompass the remaining materiel, which accounts for one third of the total weight to be moved. This section deals with the requirements for cubage in which to place this equipment; the next section deals with its weight.

Having determined the minimum number of aircraft required to move the vehicles, we must now find out if enough cubage is available in those aircraft to carry the remaining equipment. Two sources of space are available for this purpose. First, there is the space not covered by vehicles. This amounts to 28 percent (100 - 72) of the total aircraft floor area, or approximately 40 percent of the vehicle floor area (28/72). The second source is the platform area of many of the cargo vehicles. This area is roughly 30 percent of the total floor area of the vehicles. The total area available for the remaining cargo is thus equal to approximately 70 percent of the vehicle floor area.

The cubage of such equipment in the lists used in this study was approximately 1.4 times the floor area of the vehicles. (It is considerably denser.) If it were spread evenly over the available area the average stack height would thus be 2 feet (1.4/0.7). Since the minimum aircraft height considered is over 9 feet from the floor, and between 5 and 6 feet from the average truck bed, there is considerable latitude in loading this equipment. For all practical purposes, cubage per se can be disregarded in the analysis which follows.

AIRCRAFT PAYLOAD

We have determined the minimum number of aircraft required to hold the vehicles to be moved, and we have found that this number can easily hold the remaining cargo. The next step is to examine the distribution of weight per aircraft which would result if aircraft were loaded in the way the preceding Sections have described. By comparing this distribution with current designs, we can determine which constraint is more important for such aircraft—weight or floor area; more important, we can suggest desirable design criteria for future aircraft.

We will begin with the distribution of weights which results from loading the vehicles, temporarily leaving the remaining items out of the analysis. First, we must determine each vehicle's weight per square foot of floor area. For convenience, we will use the term Area Load Factor (ALF) to indicate the ratio of weight to floor area. We then calculate the distribution of these ALF's, using floor area as the frequency measure.[3] In the particular case described in this study the mean ALF was 47.6 lb/ft^2 and the standard deviation of a fitted normal curve was 33.4 lb/ft^2.

The average floor area of the vehicles studied was 108 ft^2, and we have determined that 72 percent of the floor area of any given aircraft can be covered with vehicles. Therefore, the average number of vehicles per aircraft (N) is a simple function of the aircraft floor area in square feet (A):[4]

$$N = \frac{0.72\ A}{108}.$$

If a large total number of vehicles is to be loaded, and randomly with regard to weight per square foot of floor area, it is a simple matter to estimate the distribution of the average ALF's of the individual aircraft. Each aircraft represents a sample of size (N) drawn from the population of mean 47.6 lb/ft^2 and standard deviation 33.4 lb/ft^2. The average ALF (per square foot of vehicle floor area) in a given aircraft corresponds to the sample mean, and the distribution of such average ALF's is merely the distribution of the sample means. This distribution has simple properties. The mean is, of course, the mean of the population, or 47.6 lb/ft^2. The standard deviation is simply the standard deviation of the population divided by $\sqrt{N}$ (i.e., $33.4/\sqrt{N}$).[5]

---

[3] This is the proper measure for the special case where floor-area utilization is constant. In the general case the particular measure which best represents the amount of an aircraft utilized by each vehicle should be chosen.

[4] In the general case N would have to be determined for each individual compartment size in accordance with the results of the analysis of length and width utilization.

[5] Of course this will give a somewhat optimistic estimate of the actual standard deviation, because N is merely the average number of vehicles loaded in each aircraft; since the actual number varies around N, the standard deviation of the resulting distribution will exceed the value calculated by this method. In fact, however, the error resulting from the assumption that N vehicles are loaded in each aircraft appears to be small. The actual weights loaded in the smallest aircraft under consideration--120 inches wide by 500 inches long—were computed and compared with the distribution predicted by the use of this formula. Approximately 90 percent of the aircraft contained loads within the range which the formula predicted would contain 95 percent of the cases; approximately 95 percent of the aircraft loads were within the range which the formula predicted would include 99 percent of the cases. While the error in this case was not insignificant, it appears that using the formula will not seriously bias the study, for two reasons. First, no attempt was made in the example to adjust loads so as to reduce the range of vehicle weights; such a readjustment would have made it possible to obtain a distribution much closer to that predicted by the formula. A second reason for regarding this error as acceptable is that the particular aircraft chosen for the test is the smallest considered, having a floor area of only 420 ft^2; the error for larger aircraft is likely to be considerably less important.

Thus, we can easily determine the distribution of average weight per square foot of vehicle floor area once we know the floor area of the individual aircraft. The next step is to consider the effect of loading the remaining cargo.

Figure 4 shows the cumulative distribution of weight per square foot of vehicle floor area for an aircraft of 450 ft^2 floor area. Area is expressed in terms of total vehicle floor area, the ALF is measured in lb/ft^2 of vehicle floor area; the cumulative distribution is shown by the curve OPNR. Now we must add to this weight another group of items which weigh half as much as the vehicles themselves. Since the preceding section demonstrated that there can be considerable latitude in loading these items, they will be added here so as to yield a desirable weight distribution.

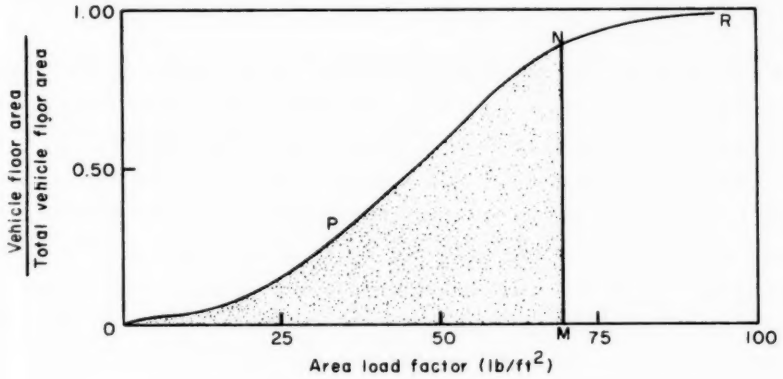


Figure 4 - Cumulative distribution of area load factors for aircraft of 450 ft^2 floor area, loaded with vehicles

What weight distribution would be the best? If the aircraft is so designed that all aircraft loaded with vehicles already contain more weight than the allowable payload, weight will be the only effective constraint on aircraft loading; the manner in which we load the remaining items is irrelevant, since they will eventually have to be reloaded in other aircraft. However, it is much more likely that only a portion of the aircraft loaded with vehicles already contain more than the allowable payload weight. In these aircraft, weight will be the effective constraint on cargo carried, but vehicle loading will be the initial constraint in the remaining aircraft. Now if additional weight is added to the former aircraft, the number of aircraft required will likewise increase—weight being the constraining dimension. However, if we load the additional weight on the latter group of aircraft (those with a low initial ALF) the number of aircraft required will not be increased—up to a point, at least—since the initial loads are below the allowable payload. Thus the preferred loading policy for this remaining equipment is quite simple: each pound should be loaded in the aircraft with the (then) smallest average ALF.

This rule can be easily performed graphically. Since the horizontal axis in Figure 4 represents ALF's measured in lb/ft^2, and the vertical axis represents floor area measured in

ft^2, an area of any given size on this graph represents some total weight. Since our task is to distribute a given total weight, we need only apply the specified rule until the required area is obtained graphically. The rule that each pound must be loaded in the aircraft with the smallest average ALF can be followed by merely moving a vertical line from left to right. The area to the left of such a line and under the curve will represent the weight added; the line is merely moved to the right until this area equals the total weight to be distributed. In Figure 4 the final position is shown by the line MN, while the shaded area OPNMO is equal to the weight of the additional cargo. The final distribution of ALF's is now given by the curve OMNR.

Thus far, we have determined the distribution of weight per square foot of vehicle floor area. We must now convert this into the weight per square foot of aircraft floor area. This is a simple operation since the vehicle floor area equals 72 percent of the aircraft floor area. By merely multiplying the ALF's already obtained by 0.72, we obtain the corresponding ALF's stated in terms of aircraft floor area.

Figure 5 represents the final distribution of weights per square foot of aircraft floor area for a cargo compartment of 450 ft^2 floor area. Consider an aircraft which can carry a payload over the relevant range equal to 52 lb/ft^2 of floor area. From Figure 5 it is evident that 90 percent of the aircraft loaded up to the floor-area constraint will not require the full payload available; for these aircraft, the limiting factor is their floor area. To determine the number of aircraft required to carry this equipment, the total vehicle floor area is divided by 0.72 and then multiplied by 0.9; the resulting figure is then divided by the floor area of the individual aircraft. Now the remaining 10 percent of the aircraft (as initially loaded) contains cargo which weighs more than the allowable payload; thus some of the equipment must be off-loaded and placed on additional aircraft.[6] Weight will be the limiting factor in loading this equipment. The total weight is shown by the shaded area in Figure 5; the total number of aircraft required to carry these items can be determined by dividing the total weight by the relevant payload of the aircraft. This procedure, although somewhat cumbersome, can provide a rather accurate estimate of the number of aircraft of a particular design required to move such a group of equipment.

Let us now summarize the ALF relationships for aircraft of various floor areas. In Figure 6, the minimum ALF in lb/ft^2 of aircraft floor area is shown by the curve labeled "0 percent," for aircraft floor areas from 400 to 1100 ft^2. It has a very slight slope and is asymptotic to 51.4 lb/ft^2. The other three curves show the values below which lie the average ALF's of 95.0, 97.5, and 99.5 percent, respectively, of the aircraft for each compartment size. The equations for these functions can be readily determined from the confidence limits of the normal distribution and the relationship among the aircraft floor area, the number of vehicles per aircraft, and the standard deviation of the relevant ALF distribution described earlier in this section.

---

[6] But it cannot be placed in any of the aircraft which are floor-area-limited since this equipment also requires floor area.

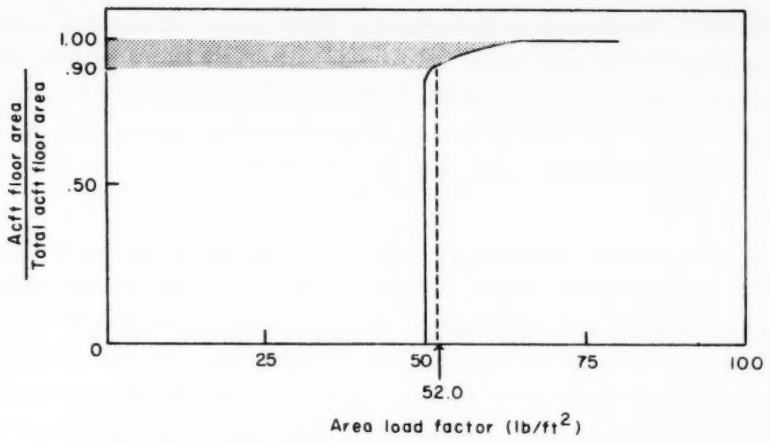


Figure 5 - Cumulative distribution of area load factors for aircraft of 450 ft^2 floor area, loaded with vehicles and other cargo

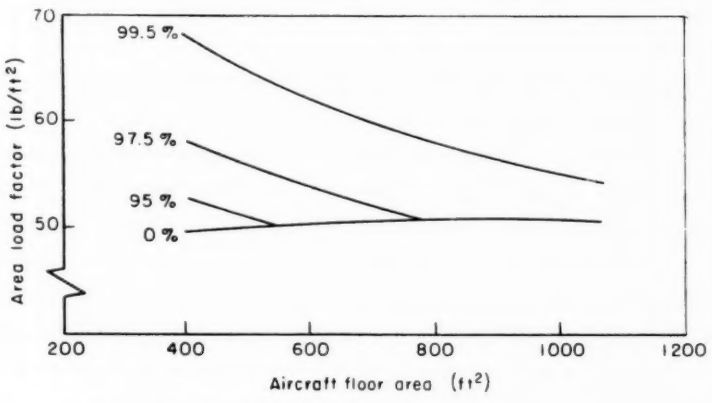


Figure 6 - Confidence limits of area load factors for various aircraft compartment sizes

One of the aims of this paper is to arrive at relationships which are simple but nonetheless take into account the more important factors to be considered in evaluating the capability of various aircraft in performing the army deployment mission. In keeping with the quest for simplicity, it seems reasonable to suggest that, at least for the larger aircraft, no great harm would be done if the distribution in Figure 6 were represented by a single line at, say, 51 lb/ft^2.[7] If this simplification is made, a straightforward rule can be adopted. Any aircraft which is designed for a maximum payload (over the relevant range) of less than 51 lb/ft^2 of its compartment floor area will be weight-limited in performing this mission; and the

[7] The slight loss of accuracy from such an approximation should be well worth the resultant gain in simplicity.

total number of aircraft required can be determined by dividing the total weight to be carried by the aircraft payload. On the other hand, any aircraft which has a payload greater than 51 lb/ft^2 of floor area will be floor-area-limited in performing this mission, and the total number of aircraft required can be determined by first dividing the total floor area of the vehicles to be moved by 0.72, and then dividing the result by the aircraft floor area.

CONCLUSIONS

As it is formulated here, the design criterion which indicates an aircraft's efficiency in performing the army deployment mission is fairly obvious. The most efficient design would provide for a payload capacity (over the relevant range) of approximately 51 lb/ft^2 of floor area. If the payload were greater than that, the payload limitation could be reduced (thus reducing the total cost of purchase and operation of the aircraft) without decreasing the amount of cargo which the aircraft could in fact carry. On the other hand, if the payload were less than 51 lb/ft^2, the compartment floor area could be reduced (again reducing the purchase and operating cost) without decreasing the amount of cargo which the aircraft could actually carry. Thus, the most efficient aircraft for this mission would have payloads (over the required range) equal to about 51 lb/ft^2 of compartment floor area.

It is important to remember that the army deployment mission is not likely to be the sole mission for any aircraft. Thus, the criterion obtained from the application of the method described here will be but one of a number of criteria to be considered in determining aircraft design. Indeed, there is no a priori reason why aircraft should be designed to accommodate existing army equipment designs and mixes. It may well be that the redesign of army force equipment, taking into account the constraints on airlift of present equipment, may be the better approach. The proper policy will depend on the extent to which changes for the purpose of this deployment will harm the efficiency of the vehicles or the aircraft in meeting their other requirements, and the importance of the deployment mission relative to those other requirements.

ACKNOWLEDGMENT

A number of members of the Mobility Project of The RAND Logistics Department listened patiently to descriptions of this study as it progressed. Many of the techniques described herein are the result of the subsequent discussions. The author is especially indebted to D. M. Fort, W. A. Niskanen, A. H. Pascal, and J. R. Summerfield.

* * *

A MODEL FOR PROCUREMENT, ALLOCATION, AND REDISTRIBUTION FOR LOW DEMAND ITEMS*†

G. Hadley

University of Chicago

and

T. M. Whitin

University of California, Berkeley

Problems of inventory control for items with extremely low demand (say less than one unit per month average demand) have received relatively little attention in the literature. However, for military supply systems it is well known that an extremely large proportion of the total number of items fall into this category. The problem of designing a control system for these low demand items is therefore one of the most critical inventory management problems faced by the National Military Establishment.

In the following, a single echelon, multidepot supply system is studied for low demand items having a stationary Poisson probability distribution for demand. Instantaneous information concerning inventory levels is assumed to be available. Procurement lead time is assumed to be constant as well as the time required for either of two available modes of redistribution. Items are ordered one at a time and decision rules are developed for allocation of new procurement, redistribution of stocks among the depots, and for determining system and depot stockage objectives in order to minimize the expected costs resulting from system and depot stockouts, cost of redistributing stocks among the depots, and costs of transportation from the source. The model is an extension of previous models in that redistribution costs and depot stockout costs are considered in the determination of the stockage objective for the system as a whole. When more than a single unit is on order, the allocation of a unit ready to be delivered is determined by the solution of a dynamic programming problem. When only a single unit is on order (and this is the one ready to be delivered), the optimal allocation procedure is reduced to allocating the unit to the depot which has the greatest probability of using it in a time period $T + (1/\lambda)$ where T is the procurement lead time and λ the system demand rate for the item.

INTRODUCTION

A supply system consisting of N depots and a central control point will be studied. It is assumed that the system operates using transactions reporting so that at any instant of time,

* Manuscript received August 26, 1960.

†Research supported by the Advanced Logistics Research Division, Bureau of Supplies and Accounts, U. S. Navy, and United Research, Inc.

the central control point knows the state of each of the N depots. The central control point has the authority to allocate procurement, to make redistribution decisions, and to determine how much stock will be carried at each of the depots. It will be supposed that the low demand item has characteristics which make it desirable to procure it one unit at a time. Demands will be assumed to occur one at a time and at random at each of the depots, the demand at one depot being independent of that at the other depots. The random variable u representing the number of units demanded at any depot j in a time interval t will be assumed to be Poisson distributed. The average demand rate λ_j at depot j is assumed to be invariant with time. Furthermore any units sent to a depot will ultimately be demanded. The procurement lead time will be imagined to be a constant T.

For this system a number of problems face the decision maker at the central control point. First there is the question of how many units should be stocked at each of the depots. For some of the depots it may be best not to stock any at all. Then there is the problem of selecting the depot to which each new unit procured will be allocated. As will be seen later, it will not always be optimal to allocate the unit to the depot where a demand initiated the order for the unit. Finally, there is the problem of deciding when to redistribute and what mode of transportation should be used. Going along with this is the similar problem of deciding upon the mode of transportation to be used in shipping from the source. A model will be developed herein which helps to answer these questions, when two modes of transportation referred to as fast and slow are allowed.

SYSTEM AND DEPOT STOCKAGE OBJECTIVES

When units are ordered one at a time as demanded, the amount of stock on hand plus on order minus backorders for the system remains constant throughout time. Denote by M the stock position of the system (on hand plus on order minus obligations). The optimal value of M is found by balancing carrying costs against the costs of stockout and redistribution. Average procurement costs (for a given mode of transportation from the source) are fixed in this model, since units are ordered one at a time, and hence do not enter into the optimization.

It will be supposed that the cost of carrying one unit in inventory for one year will be IC where C is the unit cost of the item. Furthermore it will be assumed that there exist two kinds of stockout costs, depot stockout costs and system stockout costs. The depot stockout cost will be denoted by π_j at depot j and is the cost per unit per unit time out of stock (cost of a unit year of shortage). The system stockout cost per unit year π will be assumed to be incurred in addition to the depot stockout costs if the entire system has no stock on hand. Let p(x) be the probability that the system procurement lead time demand will be x. Take λ to be the average system rate of demand for the item; then

$$\lambda = \sum_{j=1}^{N} \lambda_j$$

where as indicated previously λ_j is the average rate of demand at depot j. If the demands at the depots are independent and Poisson distributed, then p(x) also has a Poisson distribution, i.e.,

$$p(x) = \frac{(\lambda T)^x}{x!} e^{-\lambda T}$$

To introduce the approach, consider first the case where redistribution is free and instantaneous. Then there can only be system stockout costs. Furthermore, the question of how much should be stocked at each depot is irrelevant. Only the system stock position is of interest. It is found by balancing the holding costs and the system stockout costs. Recall that the system level M remains constant in time. Also within a leadtime all M units over and above the backorders will become available to meet demands. This means that the probability that w units are on hand in the system at any instant of time is simply p(M - w), i.e., is the probability that exactly M - w units are demanded in a procurement lead time. Thus, the expected on-hand inventory at any time is

$$\sum_{x=0}^{M} (M - x)\ p(x),$$

and the expected number of stockouts (backorders) is

$$\sum_{x=M+1}^{\infty} (x - M)\ p(x).$$

The expected cost per year K(M) is, then,

$$K(M) = IC \sum_{x=0}^{M} (M - x)\, p(x) + \pi \sum_{x=M+1}^{\infty} (x - M)\, p(x). \tag{1}$$

If M minimizes (1), it is necessary that $\Delta K(M) = K(M+1) - K(M) \geq 0$, and $\Delta K(M-1) < 0$. These relations show that the optimal M is the smallest M satisfying

$$\sum_{x=M+1}^{\infty} p(x) \leq \frac{IC}{IC + \pi}. \tag{2}$$

This result has been obtained previously [1]. The M obtained in this way will be called the "free and instantaneous" M since it is optimal when redistribution is free and instantaneous.

Now consider the case where redistribution is neither free nor instantaneous. In order to include the costs of redistribution and depot stockouts in the total cost expression it is necessary to say something about redistribution. The purpose of redistribution is to prevent or eliminate stockouts in the period between allocations. For the purpose of setting the system

and depot stock positions it will be assumed that a redistribution is never made to a depot until there is a backorder. The assumption is reasonable because for the low demand items under consideration, the inventory levels at most depots will frequently be either 0 or 1 and hence normally a redistribution will not be made until there is a backorder. This does not imply that the redistribution rules will never indicate the need for a redistribution until there is a backorder. It only implies that the probability of this event is small enough so that it does not have an appreciable influence on the system stock position. It will also be assumed that when there is a stockout a redistribution will be made and that the stockout lasts for the full redistribution lead time, i.e., no allocation arrives during the redistribution lead time. With these assumptions, the cost of a stockout at depot j (to be added to any system stockout costs) is

$$\hat{\pi}_j = v + \pi_j T_M, \tag{3}$$

where v is the cost of redistribution by the appropriate mode of transportation (the selection of this mode will be discussed later) and T_M is the redistribution lead time. Both v and T_M will be average values since they really depend on which depot redistributes to depot j.

Since the purpose of redistribution is to protect against stockouts in the time period between allocations, it will be necessary to compute the probability that there will be u demands at depot j in the time period between allocations. Because lead times are fixed the time period between allocations will have the same distribution as the time period between demands for the system, that is, $\lambda e^{-\lambda t}$ when demands are assumed to be Poisson. Then the marginal distribution for demand $p_j(u)$ at depot j in the time period between allocations is, for Poisson demands,

$$\hat{p}_j(u) = \int_0^\infty \frac{\lambda(\lambda_j t)^u}{u!} e^{-(\lambda + \lambda_j)t}\, dt = \left(\frac{\lambda_j}{\lambda_j + \lambda}\right)^u \left(\frac{\lambda}{\lambda_j + \lambda}\right), \quad u \geq 0 \tag{4}$$

which is the geometric distribution. Equation (4) follows immediately on integration by parts.

Denote by $d_j(\xi)$ the probability that there are ξ units on hand at depot j immediately after an allocation. The expected number of stockouts at depot j in the time period between allocations, if ξ units are on hand at the beginning of the period, is [1]

$$\sum_{u=\xi+1}^{\infty} (u - \xi)\, \hat{p}_j(u). \tag{5}$$

[1] Note that the probability of more than one demand at depot j per period will normally be small so that the expected number of stockouts will be essentially the same as the probability of one backorder, i.e.,

$$\sum_{u=\xi+1}^{\infty} (u - \xi)\, \hat{p}_j(u) \doteq \hat{p}_j(\xi+1).$$

Thus, on averaging over ξ_j, the expected cost of stockouts and redistribution per year is

$$\lambda \sum_{j=1}^{N} \pi_j \sum_{\xi} \sum_{u=\xi+1}^{\infty} (u - \xi)\, \hat{p}_j(u)\, d_j(\xi), \tag{6}$$

Since on the average there are λ periods per year.

It appears difficult to compute accurately $d_j(\xi)$, since it depends on the allocation policy, the redistribution policy and the joint distribution of the stocks at all N depots. We shall not attempt to compute the $d_j(\xi)$. Instead we proceed as follows: Let S_j be the safety stock maintained at depot j. If ρ is the average procurement lead time demand, then ρ is the average amount on order in the system, and hence

$$M = \rho + \sum_{j=1}^{N} S_j . \tag{7}$$

Now M must be an integer. It will also be assumed that the S_j are integers also, since S_j will normally be on hand at depot j. This assumption is not completely consistent with Eq. (7) since ρ may not be an integer. In order to make M an integer when the S_j are integers, ρ will be replaced by the smallest integer greater than ρ.

The S_j will be 0 or 1 for most depots for this low demand item. To simplify Eq. (6) it will be assumed that sufficient accuracy will be obtained by assuming S_j is on hand at depot j after an allocation. Then, the average redistribution cost per year becomes

$$\lambda \sum_{j=1}^{N} \pi_j \sum_{u=S_j+1}^{\infty} \left(u - S_j\right) \hat{p}_j(u). \tag{8}$$

The total yearly expected system costs, exclusive of procurement costs, are then

$$TC = K(M) + \lambda \sum_{j=1}^{N} \pi_j \sum_{u=S_j+1}^{\infty} \left(u - S_j\right) \hat{p}_j(u). \tag{9}$$

In order for any given set of S_j to minimize Eq. (9), it is necessary that

$$\Delta TC\left(S_j\right) = TC\left(S_j + 1\right) - TC\left(S_j\right) \geq 0, \quad \Delta TC\left(S_j - 1\right) < 0, \quad j = 1, \ldots, n.$$

Thus, the smallest S_j, $j = 1, \ldots, n$ is chosen so that

$$\Delta TC\left(S_j\right) = IC - (IC + \pi) \sum_{x=M+1}^{\infty} p(x) - \lambda \pi_j \sum_{u=S_j+1}^{\infty} \hat{p}_j(u) \geq 0 . \tag{10}$$

In general, each of the inequalities (Eq. (10)) depends on all of the S_j. However, if the system stockout term is negligible,[2] the inequalities become uncoupled and the j^{th} inequality depends only on S_j. In this case, the smallest S_j is chosen so that

$$\sum_{u=S_j+1}^{\infty} \hat{p}_j(u) \leq \frac{IC}{\lambda \hat{\pi}_j} , \quad j = 1, \ \ldots \ , N. \tag{11}$$

It would seem that normally the system will be decoupled because enough additional safety stocks will be needed to avoid depot and redistribution costs so that the system stockout costs will be negligible. The difference between the M computed from Eq. (11) and the "free and instantaneous" M will be called the "augmented inventory" if this quantity is positive. The augmented inventory is carried to avoid redistribution and depot stockout costs.

As the redistribution lead time increases, $\hat{\pi}_j$ increases, and hence the amount of augmented inventory required will increase. This observation points out the fact that the above system will not be suitable in certain cases. For example, if the redistribution lead time were longer than the procurement lead time, it would not make sense to hold safety stocks to avoid redistribution. Redistribution would never be used; units would always be ordered from the source. Before accepting the above model, then, it should be compared with a completely decentralized model in which there is no redistribution. In such a model, a stockage objective M_j (on hand plus on order minus backorders) is determined by minimizing the costs of carrying inventory and stockout costs for each depot individually. Thus the optimal M_j is the smallest M_j such that

$$\sum_{u=M_j+1}^{\infty} p_j(u) \leq \frac{IC}{IC + \pi_j} , \tag{12}$$

where $p_j(u)$ is the probability that u units will be demanded at depot j in the procurement lead time. In the event that the expected costs for this decentralized model are less than those for the centralized system, then the decentralized model should be used.

If the procurement lead time is long, compared to the redistribution lead time, much larger safety stocks will be required for the decentralized model than for the centralized model which allows for redistribution. It might be noted that even without introducing redistribution, the decentralized model could be improved upon by using an allocation rule which allocated units to depots which needed them most at the time the unit is ready to be shipped, rather than to the depots which ordered them.

---

[2] The reader may question whether or not there should be a system stockout term in addition to the depot stockout costs. It was included for generality. It might arise, in practice, as the cost of expedited procurement, for example.

Selection of Mode of Transportation from the Source

It will be observed that the mode of transportation used to ship units from the source to the depots influences p(u) but not $\hat{p}_j(u)$. Thus, if the system is decoupled, the S_j can be determined independently of the mode of transportation used from the source. The value of the mean lead time demand and hence M will, however, depend on the mode of transportation used. The costs of depot stockouts and redistribution are then independent of the mode of transportation used from the source; only K(M) depends on this mode of transportation. If c^r_{js} is the cost of shipping one unit from the source to depot j by mode r the optimal mode of transportation from the source is the one which minimizes

$$\sum_{j=1}^{N} c^r_{js} \lambda_j + K(M),$$

where M is the one appropriate to the mode of transportation under consideration. In the event that augmented inventory is needed for both modes of transportation, the slow mode[3] from the source will normally be preferred[4].

It is not hard to see that as demand increases, other things being equal, it does not pay to utilize the fast mode of transport from the source. The reason for this is that stockout costs and inventory carrying charges change (approximately) proportionately with the square root of lead time demand while transportation costs increase proportionately with demand.

Also other things being equal, it pays to use the rapid mode of transportation for high cost rather than low cost items since the potential savings in carrying charges are larger for high cost items. Similarly, as π increases it pays to utilize the fast mode for sufficiently high values of π.

Allocation Rules

The decision as to where to allocate each unit ordered can be made at the time the unit is ordered or up until the time the unit is ready for shipment into the system. Clearly, it is best to delay allocation as long as possible. Deciding on the allocation at the time a unit is ordered may lead to a poor decision. It will here be assumed that the allocation decision is made at the time the unit is ready to be shipped from the source.

---

[3] The terminology fast and slow for the modes of transportation does not necessarily refer to the mode of transportation used. For example, the same carrier might be used in both cases but the fast mode implies that expediting of the redistribution occurred. In any event the fast mode costs more and takes less time than the slow mode.

[4] The possibility of simultaneously using two modes in shipping from the source is not considered, i.e., the possibility of shipping to some depots by the fast mode and to others by the slow mode.

The allocation decision will be made by minimizing the expected costs. The planning horizon over which costs are to be minimized will now be discussed. Clearly, the best allocation will depend on the time until the next allocation. For example, because of the nature of the Poisson probabilities, if the next unit will not be available for a long time in the future, then the unit should perhaps be allocated to one of the higher demand depots where it will almost certainly be needed. However, if the next unit to be allocated will be available very soon it might be better to send the unit to one of the lower demand depots. The best allocation also depends on the number of units on order since this will determine the number to become available in a procurement lead time.

Any units on order at the time the decision is to be made will arrive within a time interval T (the procurement lead time. Furthermore, any units not on order cannot arrive in time T. In fact, the time before the next demand is on the average $1/\lambda$; if a unit is demanded and ordered at this time it will arrive at time $T + (1/\lambda)$. Thus, on the average, there will be no additional allocations from the time the last unit on order arrives until time $T + 1/\lambda$. Allocations are made to protect against stockouts until the next allocation. On the average the units on order will be the only units available to the system until $T + 1/\lambda$. The planning horizon will therefore be chosen to be $T + 1/\lambda$. If it is chosen to be longer than this, units not on order must be taken into account. If it is chosen to be shorter, not all available information will be used. In reality, of course, all future arrivals have an influence on the present decision. However, what happens after $T + 1/\lambda$ is not known and should have little influence on the allocation of the present unit.

A dynamic programming model will be developed to obtain the optimal allocation. This model takes into account the times of arrival of all units on order. Suppose that at the time the allocation decision is to be made there are w units on order including the one to be allocated. These units will be denoted by 1, 2, ... , w, and the times of allocation represented[5] by $t_1, t_2, \ldots, t_w$, where $t_1 = 0$ is the time at which the present allocation is to be made. Let

$$\tau_n = t_{n+1} - t_n, \; n = 1, \ldots, w - 1; \; \tau_w = T + \frac{1}{\lambda} - t_w. \tag{13}$$

Note that $\tau_w \geq 1/\lambda$. The time between t_n and t_{n+1} will be referred to as period n.

Denote by $f_n^j(x)$ the probability that exactly x units will be demanded at depot j in period n. Assume that the inventory on hand at depot j at time t_1 is L_j.

---

[5] For simplicity, the time required to ship the unit from the source is neglected and it is taken to be on hand immediately after the allocation decision is made. These times are typically one third or less of the redistribution times since the redistribution times include the administrative lead times as well.

At the time the allocation is to be made, there may be one depot with a backorder which has no redistribution on the way to meet it. There can be one backorder because if a demand which generates a backorder occurs within a time T_M of the next allocation no redistribution is made. In this case, the unit is automatically allocated to the depot with the backorder. There cannot be more than one backorder, since by assumption any others will be met by redistributions (even if they occurred within a time T_M of the allocation under consideration).

Consider now the case where there are no backorders at the time t_1. In setting up the dynamic programming problem it will be assumed that redistributions are made only when there is a backorder. It will also be assumed that all backorders are met by redistribution, so that at times $t_2, \ldots, t_w$ there are no backorders which do not have redistributions on the way to meet them. The only consequence of this assumption (which considerably simplifies the formulation) is that expected stockout costs are slightly overstated since any backorder which is met by allocation will avoid the redistribution cost and will also incur stockout costs for a time somewhat less than T_M. This error will be very small provided that T_M is not a large fraction of the average time between procurements.

In the dynamic programming problem the costs of stockout and redistribution are minimized. Carrying costs need not be included since each unit will be carried somewhere in the system. Transportation costs are not included either since on the average they are incurred anyway.

Take y_n^j to be the number of units allocated to depot j at time t_n. Since only one unit is to be allocated at a given time it must be true that

$$y_n^j = 0 \text{ or } 1 \text{ and } \sum_{j=1}^{N} y_n^j = 1. \tag{14}$$

Since the demands at the various depots are assumed to be independent, the probability that $\underset{\sim}{x} = (x_n^1, \ldots, x_n^N)$ is a vector of the demands at the depots in period n is

$$f_n(\underset{\sim}{x}) = \prod_{j=1}^{N} f_n^j\left(x_j\right). \tag{15}$$

If the stock at depot j is ξ_n^j at time t_n before the allocation of unit n, then the stock at depot j is

$$\xi_{n+1}^j = \xi_n^j + y_n^j - x_n^j > 0 \text{ or } \xi_{n+1}^j = 0, \tag{16}$$

because backorders are eliminated by redistribution.

Define $T_n(\xi_1, \ldots, \xi_n)$ to be the minimum cost for periods n through w when an optimal policy is followed and when the stock at depot j at the beginning of period n (before the allocation) is ξ_j. Then the recurrence relations have the form

$$T_n\left(\xi_1, \cdots, \xi_N\right) = \min_{\underset{\sim}{y}_n} \left\{ \sum_{j=1}^{N} \hat{\pi}_j \sum_{u_n^j = \xi_j + y_n^j + 1}^{\infty} \left(u_n^j - y_n^j - \xi_j\right) f_n^j\left(u_n^j\right) \right.$$

$$\left. + \sum_{\text{all } \underset{\sim}{x}_n} f_n(\underset{\sim}{x})\ T_{n+1}\left(\left[\xi_1 + y_n^1 - x_n^1\right], \cdots, \left[\xi_n + y_n^n - x_n^n\right]\right)\right\}, \tag{17}$$

where

$$\left(\xi_j + y_n^j - x_n^j\right) = \begin{cases} \xi_j + y_n^j - x_n^j & \text{if } \xi_j + y_n^j - x_n^j > 0 \\ 0 & \text{if } \xi_j + y_n^j - x_n^j \leq 0, \end{cases}$$

and

$$\underset{\sim}{y}_n = \left(y_n^1, \cdots, y_n^N\right).$$

Also

$$T_w\left(\xi_1, \cdots, \xi_N\right) = \min_{\underset{\sim}{y}_w} \sum_{j=1}^{N} \hat{\pi}_j \sum_{u_w^j = \xi_j + y_w^j + 1}^{\infty} \left(u_w^j - y_w^j - \xi_j\right) f_w^j\left(u_w^j\right). \tag{18}$$

The minimum cost is then $T_1\left(L_1, \cdots, L_N\right)$. The values y_1^j tell what the optimal allocation should be.

The computational procedure is to compute first $T_w\left(\xi_1, \cdots, \xi_N\right)$ and tabulate it for all the possible $\xi_1, \cdots, \xi_N$. Then, using T_w, compute T_{w-1} and tabulate it for all possible $\xi_1 \cdots, \xi_N$, etc., until $T_1\left(L_1, \cdots, L_N\right)$ is found along with the optimal y_1^j. If the ξ_j or u_j could vary over a wide range, it would be quite impossible to solve a dynamic programming problem of this sort even with the largest computer available today. It is only because the assumption is being made that demands are very low that anything can be done. For almost all the depots, the range of variation of ξ_j will be 0 or 1, although it is possible to allow a small number of depots to have ξ_j go up to 2 or 3. Furthermore, for almost all the depots $x_j = 0$ or 1. The probability that $x_j > 1$ will usually be negligible. In addition $\sum x_j$ will normally take on only the values 0, 1, 2. The probability of a system demand of more than 2 before the next allocation will be small since on the average, the time between allocations is the average time between system demands.

With the above simplifications, it seems quite possible to solve a two stage problem with about 10 depots on a high speed computer in something like 30 seconds or less. It seems unlikely that it would be worthwhile in practice to ever try to include more than two stages.

This should give all the accuracy needed. Note that in Eq. (17) with the simplifications suggested

$$\sum_{\text{all } \underline{x}} = \Sigma\left(\Sigma x_j = 0\right) + \Sigma\left(\Sigma x_j = 1\right) + \Sigma\left(\Sigma x_j = 2\right).$$

For example, $\Sigma(\Sigma x_j = 1)$ is the sum over all x_j such that one and only one x_j is unity and all other x_j vanish. Furthermore in $\Sigma x_j = 2$, all possibilities need not be considered, because the probability of two demands is negligible at most depots. There can still exist quite a few possibilities for the set of ξ_j since if there were 10 depots and each depot could have $\xi_j = 0$ or 1 there would be $2^{10} \sim 1000$ possibilities for $(\xi_1, \cdots, \xi_N)$.

In the event that the unit to be allocated is the only one on order, a tremendous simplification can be introduced into the dynamic programming problem since it reduces to a one stage problem. The single stage extends over the entire planning horizon $T + 1/\lambda$. Denote $f^j_{T + 1/\lambda}(u)$ by $p_j^*(u)$. Then the problem reduces to that of minimizing

$$z = \sum_{j=1}^{N} \left[\hat{\pi}_j \sum_{u_j = L_j + y_j + 1}^{\infty} \left(u_j - L_j - y_j\right) p_j^*\left(u_j\right) \right], \tag{19}$$

where

$$y_j = 0 \text{ or } 1 \text{ and } \Sigma y_j = 1.$$

Suppose $y_1^*, \cdots, y_N^*$ constitute an optimal set of y_j, that is, a set of y_j which minimize Eq. (19). Let z^* be the optimal value of z. Assume that $y_k^* = 1$ and all other $y_j^* = 0$. If the unit were allocated to depot $i \neq k$, then the value of z for this allocation must satisfy $z - z^* \geq 0$. However,

$$z - z^* = \hat{\pi}_k \sum_{u_k = L_k + 1}^{\infty} p_k^*(u_k) - \hat{\pi}_i \sum_{u_i = L_i + 1}^{\infty} p_i^*(u_i) \geq 0. \tag{20}$$

The optimal allocation of the unit is then to the depot with the largest value of

$$\hat{\pi}_k \sum_{u = L_k + 1}^{\infty} p_k^*(u).$$

In the event that the $\hat{\pi}_k$ are all equal, the result is that the unit is allocated to the depot which has the greatest probability of using it in the time period $T + 1/\lambda$, which depot may or may not be the depot whose demand originated the order for the unit to be allocated.

For many low cost items, it will not be worthwhile to use the dynamic programming model to make the allocation. For some cases, it will be sufficient simply to allocate the unit to the depot which has the greatest probability of using it in the time period until the next allocation.

In the initial provisioning of the depots, each depot will receive the safety stock S_j computed using the rules derived earlier. If the amount which is normally on order is allocated to the depots, the simplest thing to do (which should also be quite satisfactory) is to allocate the units one at a time to the depot with the greatest probability of using it in the time $T + 1/\lambda$.

Redistribution

It remains to consider the problem of redistribution. Since the on hand inventory levels will generally be very low (0 or 1 usually), it will be desirable to consider whether or not there should be a redistribution each time there is a demand in the system. The purpose of a redistribution is to protect against or reduce stockout costs in the time interval until the next allocation or in the time interval $T + 1/\lambda$ if nothing is on order.

While it is permissible (provided that the redistribution time is substantially less than the average time between allocations) in setting the stockage objectives at the depots and in making allocations to assume that a backorder will be satisfied by a redistribution and that the cost of a redistribution to j will be $\hat{\pi}_j$, these assumptions cannot generally be made when a redistribution is being considered. Recall that a redistribution is to be considered each time there is a demand in the system. If the redistribution time is greater than the time until the next allocation, no redistribution is considered. In such a case, if the demand caused a backorder, the next unit to arrive will automatically be allocated to the depot with the backorder, provided that another backorder does not occur before the allocation time.

Consider now the case where a second backorder occurs in a time period T_M before a unit is to be allocated. At the point in time when the second backorder occurs, the next unit to be allocated is tentatively allocated to one or the other of the backorders by minimizing stockout and transportation costs under the assumption that a redistribution will be made to the other depot. After the allocation is tentatively made, a redistribution is considered to the remaining depot with a backorder. The time period to be used here is the time until the second unit or order is ready to be allocated or a time $t + T + 1/\lambda$ if only one unit is an order, where $t < T_M$ is the time until the unit on order is ready to be delivered. Of course, if a third backorder occurs in the time period T_M before the arrival of the first unit on order, the tentative allocation is reexamined, and a redistribution is then considered to the depot not to receive the allocation.

When a redistribution is considered, the above discussion has shown that if $t = \tau + T_M$ is the time until the next allocation, then $\tau > 0$. The next allocation here means the next one that is not tentatively already allocated; it may in certain cases be the time of arrival of the second unit on order. The only depot considered as a possible receiver of a redistribution is, of course, the depot where the demand occurred. There are two cases to be examined. The first is that where the demand caused a backorder at the depot j where it occurred. The other is that where it did not cause a backorder.

The most important case is that where the demand at depot j generates a backorder there. In this case the mode of transportation for redistribution can be selected at the outset. It is the mode which minimizes the sum of transportation and stockout costs i.e., minimizes $\hat{\pi}_j = v + \pi_j T_M$. It will almost always be true that the fast mode of transportation should be used in this case. The depot i which is considering a redistribution to j must have at least one unit on hand; let L_i be the number of units on hand at i. It will be assumed in making the computations that no units will later be redistributed to i unless there is a backorder If a backorder occurs at i before the next allocation, it will be assumed that a redistribution will be made by the mode of transportation decided upon above at a cost $\hat{\pi}_i$ if the redistribution occurs at a time at least T_M before the arrival of the next allocation. If the time of the backorder is less than T_M before the arrival of the next allocation it will be assumed that the unit will be allocated to i and the stockout costs will be $\pi_i t$ where $t \leq T_M$ is the time until the next allocation. When more than a single backorder occurs at i before the next allocation it is supposed that all after the first will be met by redistribution regardless of when they occur, except in the case where more than L_i demands occur in the time period of length τ. In this case, the first demand, if any, occuring within a time $t \leq T_M$ of the next allocation can be met by allocation rather than redistribution.

The expected costs of stockout and redistribution for depots i, j will now be computed both for the case where i makes a redistribution to j and the case where i does not make a redistribution to j. The expected costs at the other depots have no influence on the decision.[6] If i does not ship to j, the cost at j will be $\pi_j (\tau + T_M)$ since there is a backorder at j. Costs arising from additional backorders at j occuring later have no influence on the decision because the redistribution does not modify then. At depot i stockout costs will be incurred if more than L_i units are demanded in the time $\tau + T_M$. The cost of the first backorder (demand $L_i + 1$) depends on whether it occurs at a time $t > T_M$ before the next allocation or a time $t \leq T_M$ before the next allocation. It will be $\hat{\pi}_i$ if $t > T_M$ and $\pi_i t$ if $t < T_M$. The cost of any backorders after the first (demands $L_i + 2$, etc.) will be $\hat{\pi}_i$ except as noted previously, in the case where more than L_i units are demanded in the time period of length τ.

The probability that demand $L_i + 1$ occurs in the interval 0 to τ is

$$\sum_{u=L_i+1}^{\infty} f_{\tau}^{i}(u)$$

---

[6] In actuality, depots i, j cannot be completely isolated from the rest of the system. There are second order interactions. For example, if a backorder occurs at i in a time $t < T_M$ before the arrival of the next allocation, then whether or not this backorder is met by allocation or redistribution depends on whether or not some other depot has also had a backorder in the time period of length T_M before the next allocation. These second order effects are ignored here since they have only a very slight influence on the expected costs.

where $f^i_\tau(u)$ is the Poisson probability that exactly u units are demanded at depot i in a time period of length τ. The probability that demand $L_i + 1$ occurs between t and $t + dt$, $0 \leq t \leq T_M$ (t being time until the next allocation) is $\lambda_i f^i_{\tau+t}(L_i)\,dt$. Thus the expected costs at depots i, j until the next allocation if no redistribution is made are[7]

$$D = \pi_j(\tau + T_M) + \hat{\pi}_i \sum_{u=L_i+2}^{\infty} (u - L_i - 1) f^i_{\tau+T_M}(u) + \hat{\pi}_i \sum_{u=L_i+1}^{\infty} f^i_\tau(u) + \lambda_i \pi_i \int_0^{T_M} t\, f^i_{\tau+t}(L_i)\,dt$$

$$+ \lambda_i \pi_i \left[\sum_{u=L_i+1}^{\infty} f^i_\tau(u)\right] \int_0^{T_M} t\, f^i_t(0)\,dt, \tag{21}$$

where

$$\int_0^{T_M} t\, f^i_{t+\tau}(L_i)\,dt = \frac{L_i+1}{\lambda_i 2}\left[\sum_{u=L_i+2}^{\infty} f^i_{\tau+T_M}(u) - \sum_{u=L_i+2}^{\infty} f^i_\tau(u)\right]$$

$$-\frac{\tau}{\lambda_i}\left[\sum_{u=L_i+1}^{\infty} f^i_{\tau+T_M}(u) - \sum_{u=L_i+1}^{\infty} f^i_\tau(u)\right]. \tag{22}$$

In particular,

$$\int_0^{T_M} t\, f^i_t(0)dt = \int_0^{T_M} t\, e^{-\lambda_i t} dt = \frac{1}{\lambda_i 2}\left[1 - e^{-\lambda_i T_M}\left(1 + \lambda_i T_M\right)\right]. \tag{23}$$

If a redistribution is made to j, the expected costs are

$$D^* = \hat{\pi}_j + \hat{\pi}_i \sum_{u=L_i+1}^{\infty} (u - L_i) f^i_{\tau+T_M} + \hat{\pi}_i \sum_{u=L_i}^{\infty} f^i_\tau(u) + \lambda_i \pi_i \int_0^{T_M} t\, f^i_{\tau+t}(L_i - 1)\,dt$$

$$+ \lambda_i \pi_i \left[\sum_{u=L_i}^{\infty} f^i_\tau(u)\right] \int_0^{T_M} t\, f^i_t(0)\,dt. \tag{24}$$

---

[7] The term

$$\sum_{u=L_i+2}^{\infty} (u - L_i - 1) f^i_{\tau+T_M}(u)$$

is an approximation for the sum of three terms, the first giving the expected backorders from demands $L_i + 2, \ldots,$ in the period of length τ, the second giving the expected backorders from demands $L_i + 2, \ldots,$ in the period of length T_M given that demand $L_i + 1$ occurs in this period, and the third giving the expected number of backorders above one in the period of length T_M given that more than L_i occur in period of length τ, given that one occurs in the period of length T_M. This should be quite satisfactory, since the probability of more than two demands in a time $\tau + T_M$ will be small.

Expected costs will be reduced by making a redistribution if $\Delta D = D^* - D < 0$. However,

$$\Delta D = \hat{\pi}_j - \pi_j t + \hat{\pi}_i \sum_{u=L_i+1}^{\infty} f^i_{\tau+T_M}(u) + \hat{\pi}_i f^i_\tau(L_i) + \frac{\pi_i L_i}{\lambda_i}\left[f^i_{\tau+T_M}(L_i+1) - f^i_\tau(L_i+1)\right]$$

$$- \frac{\pi_i}{\lambda_i}\left[\sum_{u=L_i+2}^{\infty} f^i_{\tau+T_M}(u) - \sum_{u=L_i+2}^{\infty} f^i_\tau(u)\right] - \pi_i \tau\left[f^i_{\tau+T_M}(L_i) - f_\tau(L_i)\right]$$

$$+ \pi_i \lambda_i f^i_\tau(L_i) \int_0^{T_M} t\, f^i_t(0)\, dt. \tag{25}$$

For each depot i with $L_i > 0$, ΔD is computed. If every $\Delta D \geq 0$, no redistribution should be made. When one or more $\Delta D < 0$, a redistribution should be made, and the depot to make the redistribution is the one with the most negative ΔD.

The end effect terms in Eq. (24) which arise because a backorder may be met by allocation rather than redistribution will yield expected costs significantly different from those obtained if it is assumed that all backorders are met by redistribution only when $L_i = 1$ (the shipping depot has only a single unit) and the time at which the backorder occurs is close to the time T_M before the next allocation. These differences can, in general, be important however, since for the case under consideration it will usually be true that $L_i = 1$.

It remains to consider the case where the demand at depot j does not generate a backorder. Thus, after the demand, $L_j \geq 0$. This case will be studied in two parts: (a) that where $L_j = 0$ after the demand and (b) that where $L_j > 0$ after the demand. Part (a) will be examined first. Not only must it be decided whether or not to make a redistribution, but in addition the mode of transportation must be selected if a redistribution is to be made. By using the fast mode of transportation, the expected stockout costs in the time until the redistribution arrives can be reduced; however the fast mode is more expensive.

The expected costs at depot j and the prospective shipping depot i will now be found for the case where i does not make a redistribution. As above, the time until the next allocation will be written $t = \tau + T_M$. For depot j (which now has no units on hand), the stockout costs incurred if there is more than a single demand in the time period t will not be considered since a redistribution could not influence them. It will be assumed that if there is a backorder at a time t before the next allocation $T_M < t < \tau + T_M$, it will be met by redistribution at a cost $\hat{\pi}_j$, and if $0 < t \leq T_M$, it will be met by allocation at a cost $\pi_j t$. For depot i, it will be assumed that any backorder will be met by redistribution at a cost $\hat{\pi}_i$. The end effect where a backorder can be met at i by allocation will be ignored, since two demands must

occur to generate such a backorder, thus making its effect small. Therefore, the expected costs for i, j until the next allocation are:

$$(26)\quad D = \hat{\pi}_i \sum_{u=L_i+1}^{\infty} (u-L_i)\, f^i_{\tau+T_M}(u) + \hat{\pi}_j \sum_{u=1}^{\infty} f^j_{\tau}(u) + \lambda_j \pi_j \int_0^{T_M} t\, f_{\tau+t}(0)\, dt.$$

The expected costs if i does make a redistribution will be found next. For depot j, it follows that stockout costs are incurred only if there is a demand before the arrival of the redistribution, since stockouts from more than one demand need not be included. Now, the end effect will be included for depot i rather than j since typically, i will have only a single unit on hand before it makes the redistribution. The expected costs at i, j if a redistribution is made are then

$$D^* = \hat{\pi}_i \sum_{u=L_i+1}^{\infty} (u-L_i)\, f^i_{\tau+T_M}(u) + \hat{\pi}_i \sum_{u=L_i}^{\infty} f^i_{\tau}(u) + \lambda_i \pi_i \int_0^{T_M} t\, f_{\tau+t}(L_i-1)\, dt$$

$$(27)\quad + \lambda_i \pi_i \left[\sum_{u=L_i}^{\infty} f^i_{\tau}(u)\right] \int_0^{T_M} t\, f_t(0)\, dt + \pi_j^*,$$

where π_j^* is the expected stockout cost at j before the arrival of the redistribution.

$$\pi_j^* = \pi_j \int_0^{T_M} \lambda_j \left(T_M - t\right) e^{-\lambda_j t}\, dt = \pi_j \left(T_M - \frac{1}{\lambda_j}\right)\left(1 - e^{-\lambda_j T_M}\right)$$

$$(28)\quad + \pi_j T_M e^{-\lambda_j T_M}.$$

If $\Delta D = D^* - D < 0$, expected costs could be reduced by making the redistribution. The ΔD values are computed for each mode of transportation, and if all $\Delta D \geq 0$ no redistribution should be made; if one or more $\Delta D < 0$, then a redistribution should be made by the depot and mode of transportation which minimize ΔD.

Finally, part (b) will be studied. In this case, there are $L_j \geq 1$ units on hand at j after the demand. Only on very rare occasions would a redistribution be made to j in this case. It would be done, only if j was one of the relatively high demand depots and the time until the arrival of the next allocation was considerably greater than the average time between allocations. Here the end effect can be ignored and the cost of a backorder at j can be taken to be $\hat{\pi}_j$. Now,

$$(29)\quad \Delta D = v + \hat{\pi}_i \sum_{u=L_i}^{\infty} f^i_{\tau+T_M}(u) - \hat{\pi}_j \sum_{u=L_j+1}^{\infty} f^j_{\tau+T_M}(u).$$

If $\Delta D \geq 0$ for all $i \neq j$, a redistribution should not be made. If $\Delta D < 0$ for one or more i, a redistribution should be made by the depot which minimizes ΔD. Here, it will almost always be true that if a redistribution should be made, the slow mode of redistribution should be used.

This completes the discussion of redistribution. A high-speed computer could go through this redistribution routine each time an item is demanded in a couple of seconds or less. However, the details could in many cases be carried out quite well by hand without going through all the details developed herein (just described), since often it will be fairly obvious which depot should redistribute to the needy depot, if no redistribution is made until there is a backorder.

A Numerical Example - A system consisting of three depots will be used for the example. The following data are assumed:

$IC = \$200$

$\pi = \$100,000$

$\hat{\pi}_j = \$2000$ for $j = 1, 2, 3$.

$N = 3$

$\lambda = 3$, $\lambda_1 = 2$, $\lambda_2 = 0.6$, $\lambda_3 = 0.4$ per year.

Procurement Lead Time = 0.25 year

Redistribution Lead Time = 0.02 year

$\rho = 0.75$, (round to 1).

Under the assumption that system stockout costs are negligible, the "uncoupled" model can be used. Inequalities (Eq. (11)) with the $\hat{p}_j(u)$ of Eq. (4) are used to determine the safety stocks. The $\hat{p}_j(u)$ are tabulated in Table 1.

TABLE 1

	Depot 1	Depot 2	Depot 3
$\hat{p}_j(0)$	0.6	0.833	0.883
$\hat{p}_j(1)$	0.24	0.1388	0.1039
$\hat{p}_j(2)$	0.096	0.0231	0.0122
$\hat{p}_j(3)$	0.00384	0.00386	0.00143
$\hat{p}_j(4)$	0.00154	0.00064	0.00017

Inequalities (Eq. (11)) result in optimal values[7], $S_1 = 2$; $S_2 = 1$; $S_3 = 1$. For $\rho = 0.75$ (rounded to 1), Eq. (7) gives a system stockage objective, (on hand plus on order minus backorders) of 5. Note that the uncoupling assumption was quite realistic since the probability of system backorders is only about 0.0001 per procurement lead time.

[7] It might be noted that the same safety stocks would be obtained if $\hat{p}_j(u)$ was taken to be Poisson distributed with mean λ_j/λ, i.e., if the demand in the meantime between allocations was used.

To illustrate the nature of the dynamic programing allocation rule, it is assumed that there are two units on order, and that at the time the first unit is ready to be allocated ($t_1 = 0$), the on hand levels are $\xi_1 = 1$, $\xi_2 = 2$, $\xi_3 = 0$. The second unit will be ready for allocation after 0.3 years elapse, ($t_2 = 0.3$). The lengths of the two periods under consideration are therefore $\tau_1 = 0.3$, $\tau_2 = 0.28$ (See Eq. (13).).

At the start of the last period, the possible values of $\underset{\sim}{\xi} = (\xi_1, \xi_2, \xi_3)$ are given in Table 2 along with the expected number of stockouts in the last period. The optimal allocation of the unit allocated at $t_2 = 0.3$, and the expected costs in the last period.[8]

TABLE 2

ξ_1	1	2	0	0	1	1	0	0	0	1	1	0	2	2	1
ξ_2	2	2	2	2	1	0	1	0	0	1	0	1	1	0	2
ξ_3	1	0	1	0	1	1	1	1	0	0	0	0	0	0	0
Optimal depot (for allocation at $t_2 = 0.3$)	1	3	1	1	1	2	1	1	1	1	2	1	3	2	1
Expected Stockouts	0.025	0.025	0.145	0.25	0.035	0.155	0.155	0.305	0.41	0.14	0.26	0.26	0.035	0.14	0.13
Expected Cost	50	50	290	500	70	310	310	610	820	280	520	520	70	280	260

TABLE 3

(a) $\underset{\sim}{\xi}$	(0,2,0)	(1,1,0)	(0,1,1)	(0,1,1)	(1,0,1)	(1,1,0)	(0,2,1)	(1,1,1)	(1,2,0)	(1,2,1)
(b) Period 1 Demand	(1,0,1)	(0,1,1)	(1,1,0)	(2,0,0)	(0,2,0)	(0,0,2)	(1,0,0)	(0,1,0)	(0,0,1)	(0,0,0)
(c) Probability	0.0294	0.00875	0.0495	0.0745	0.00487	0.00277	0.246	0.073	0.049	0.41
(d) No. first period Redistributions	0	0	0	1	0	1	0	0	0	0
(e) Cost of first period redistributions	0	0	0	2000	0	2000	0	0	0	0

[8] To simplify computations some possible $\underset{\sim}{\xi}$ with very low probabilities have been neglected here. Specifically, cases with system demands in a single time period of more than two units are ignored.

TABLE 3 (Cont'd)

(f) Expected minimum cost in period 2	280	280	280	500	500	280	260	140	140	50
(g) Costs x Prob.	14.70	2.44	15.30	172	1.51	6.30	71.50	5.12	12.70	20.50

TABLE 4

(a) $\underset{\sim}{\xi}$	(1,1,0)	(1,1,0)	(1,1,0)	(0,2,0)	(0,2,0)	(1,1,0)	(1,2,0)	(2,1,0)	(2,1,0)	(2,2,0)
(b) Period 1 Demand	(1,0,1)	(0,1,1)	(1,1,0)	(2,0,0)	(0,2,0)	(0,0,2)	(1,0,0)	(0,1,0)	(0,0,1)	(0,0,0)
(c) Probability	0.0294	0.00875	0.0495	0.0745	0.00487	0.00277	0.246	0.073	0.049	0.41
(d) No. first period redistributions	1	1	0	0	0	2	0	0	1	0
(e) Cost of first period redistributions	2000	2000	0	0	0	4000	0	0	2000	0
(f) Expected minimum cost in period 2	280	280	280	500	500	280	260	140	140	50
(g) Costs x Prob.	67.0	19.90	13.85	37.25	2.43	11.85	64.0	10.20	105.0	20.50

It remains to calculate the optimal allocation at $t_1 = 0$, given that the initial $\underset{\sim}{\xi}$ is (1,2,0). Since depot 2 is clearly overstocked relative to the other two depots, for simplicity only allocation to depots 1 and 3 is considered here. Table 3 gives (a) the values of $\underset{\sim}{\xi}$ at $t_2 = 0.3$ if allocation of the initial unit (t = 0) is to depot 3 when the period one demand is that given in row (b). The table also presents: (c) probability of occurrence; (d) the number of distributions in period 1; (e) the cost of period 1 redistributions; (f) the costs of period 2 for the specified $\underset{\sim}{\xi}$ vector; (g) the cost for periods 1 and 2 weighted by the appropriate probability. The sum of the last row gives total expected cost for the two periods of $322.07.

Table 4 provides the same data as Table 3 for the case where the unit allocated at $t_1 = 0$ is allocated to depot 1, i.e., the inventory after the first allocation is (2, 2, 0). The expected total cost is $351.98, about $29 more than that resulting from allocating the unit available at $t_1 = 0$ to depot 3. Hence, optimal policy is to allocate the first unit to depot 3.

REFERENCE

[1] T. M. Whitin and J. W. T. Youngs, "A Method for Calculating Optimal Inventory Levels and Delivery Time," Naval Research Logistics Quarterly, September 1955, pp. 157-173.

* * *

ON MINIMAX SURVEILLANCE SCHEDULES*†

C. Derman
Columbia University

INTRODUCTION

Consider, as a unit, an operating system, a component of such a system, or a lot of material maintained in storage for use at some future unknown time. Suppose the unit under consideration deteriorates and that after a random period of time the unit is no longer capable of functioning properly. We shall refer to this length of time as the life of the unit. In many cases there are estimable costs associated with the continuance in service (storage) of a unit which has failed. If the unit is a production system, the costs are associated with the amount of defective product produced; if it is material in storage (e.g., certain kinds of missile fuels), the costs are derived from considering the various implications of using, unknowingly, the unserviceable material; ... and the like. Thus, in such cases, it is usual to inspect the unit from time to time in order to determine its condition. However, since the inspections are often costly, some consideration should be given to the scheduling of the inspections.

We treat this surveillance-scheduling problem assuming that the unit remains in service (storage) until failure is detected by an inspection or it reaches some stipulated time T in service. We assume that a failure is detected by an inspection with probability p ($p > 0$).

We do not assume any knowledge concerning the distribution function of the life of the unit. Our aim is to derive the schedule which minimizes the maximum possible expected cost associated with the inspections and failure of the unit.

Of course, if the distribution is known, then our minimax criterion is not meaningful. In fact, our particular formulation of the problem[1] was suggested by the one given by Barlow and Hunter [1] . In their formulation the distribution function F is assumed known and $p = 1$, and they derive the schedule which minimizes the expected cost associated with F. Roeloffs (in current work to appear in his doctoral dissertation) has obtained results relating to optimal schedules when partial knowledge concerning F is available.

THE FORMAL PROBLEM

Let Y, with unknown distribution function F, denote the life of the unit. By a surveillance schedule consisting of n observations we mean a set of n numbers $x_i (i = 1, \ldots, n)$ with

---

*Manuscript received November 18, 1960.

†Research sponsored by the Office of Naval Research.

[1]Our interest in this problem area originally grew out of conversations with Dorothy M. Gilford of ONR and representatives of the Bureau of Weapons.

$o < x_1 < \ldots < x_n < T$; x_i denotes the time at which the i^{th} inspection of the unit is made; T denotes the maximum length of time a unit is kept in service (storage). For a given $x = (x_1, \ldots, x_n)$ there is a function $G_x(Y)$ which is the cost associated with the unit having life Y. The expected cost for a given distribution function F is

$$E_F G = \int_0^\infty G_x(t)dF(t).$$

Since F is unknown, we are concerned with

$$\mu_x = \sup_F E_F G,$$

where the supremum is taken over all distribution functions of nonnegative random variables. The problem is to find that $x = x^*$ such that

$$\mu = \min_x \mu_x = \mu_{x^*}.$$

THE MINIMAX SURVEILLANCE SCHEDULE

We assume that the cost associated with the unit is

$$cN + vZ,$$

where $c > o$ is the cost of an inspection, N is the number of inspections on the unit, $v > o$ denotes the cost per unit time attributable to the unit during the interval of time between its failure and its detection, and Z denotes the length of time between failure and detection. On noting that, for any given Y, the cost is dependent on the number of subsequent inspections needed to detect the failure and on taking the conditional expectation of the cost with respect to this aspect of the existing randomness, we have that $G_x(Y)$ is of the form,

$$\begin{aligned} G_x(Y) &= g(x_k, \ldots, x_n, Y) \\ &= p(k+1)c + pv(x_{k+1} - Y) \\ &\quad + p\sum_{i=k+1}^{n-1} (1-p)^{i-k}\left\{(i+1)c + v(x_{i+1} - Y)\right\} \\ &\quad + (1-p)^{n-k}\left\{nc + v(T-Y)\right\}, \qquad x_k < Y \le x_{k+1};\ k = 0, \ldots, n-1, \\ &= nc + v(T-Y), \qquad x_n < Y \le T, \\ &= nc, \qquad T < Y, \end{aligned} \tag{1}$$

where $x_0 = 0$.

It is easily seen that for any distribution function F of a positive random variable,

$$E_F G_x(Y) = \sum_{k=0}^{n} \int_{x_k}^{x_{k+1}} g(x_k, \ldots, x_n, t)\, dF(t) + nc \int_T^{\infty} dF(t)$$

$$\leq \sum_{k=0}^{n} g(x_k, \ldots, x_n, x_k + 0)\left(F(x_{k+1}) - F(x_k)\right) + nc(1-F(T))$$

$$\leq \max\ (g_{0,x}, \cdots, g_{n,x})$$

$$= \mu_x,$$

where $g_{ix} = g(x_i, \ldots, x_n, x_i + 0)$, $i = 0, \ldots, n$,

and $x_{n+1} = T$.

Hence, our problem is to find that schedule $x = x^*$ such that

$$\mu_{x^*} = \min_x \mu_x = \min_x \max\left(g_{o,x}, \cdots, g_{n,x}\right).$$

The result of this article is as follows:

The minimax schedule x^* is given by

$$x_i = ip\left[\frac{T}{np+1} + \frac{c}{2v}\left(\frac{n[(n+1)p+2]}{np+1} - (i+1)\right)\right], \quad i = o, 1, \ldots, n, \tag{2}$$

where n is the largest integer such that

$$cp^2n^2 + cp(2-p)n + 2(c - pvT) \leq o. \tag{3}$$

Proof: Let N be such that $N(N-1) \leq 2T\ v/cp$ and $n(n-1) > 2T\ v/cp$, if $n > N$. Denote by C_n the class of schedules of n inspections, and let

$$\tilde{C}_N = \bigcup_{i=o}^{N} C_i\ .$$

Define the schedule $x = x(n)$ as that schedule $x \epsilon C_n$ satisfying

$$g_{0x(n)} = g_{1x(n)} = \cdots = g_{nx(n)} = \mu_{x(n)}, \quad n \leq N. \tag{4}$$

Let C* denote the class of schedules x(O), ..., x(N). The proof will proceed as follows. (a) We show that if $x \in C_n$ for $n > N$, then $\mu_x \geq \mu_{x(N)}$. Therefore if x^* exists, it is in $\tilde{C}_N$. (b) We show that x^* exists and is a member of C*. (c) We show that Eq. (3) allows us to identify the member of C* which is x^*.

Proof of (a). In order to exhibit x(n) explicitly, note that for any schedule $x \in C_n$, on rearranging terms,

$$g_{kx} = c\left\{\sum_{i=k}^{n-1} (i+1)p(1-p)^{i-k} + n(1-p)^{n-k}\right\} + v\left\{\sum_{i=k}^{n} \Delta_i (1-p)^{i-k}\right\}, \quad k=0, \ldots, n-1, \tag{5}$$

$$g_{nx} = cn + v\Delta_n,$$

where $\Delta_i = x_{i+1} - x_i$, $i = 0, 1, \ldots, n$. From (5) we obtain,

$$(1-p)g_{kx} + v\Delta_{k-1} + ckp = g_{(k-1)x}, \quad k = 1, \ldots, n. \tag{6}$$

On solving (4) in (6), we get $x_i (i=1, \ldots, n)$ given by (2), and

$$\mu_{x(n)} = \frac{vT + \frac{nc}{2}\left\{(n+1)p+2\right\}}{np+1}, \quad n = 0, \ldots, N.$$

Because of the constraints $\Delta_i > 0$ $(i=0, 1, \ldots, n)$, x(n) is not defined if $n > N$; for, if so, $\Delta_n \leq o$.

Now, let x be any schedule in C_n where $n > N$. Trivially, $\mu_x \geq cn$. Hence,

$$\mu_x - \mu_{x(N)} \geq cn - \frac{vT + \frac{Nc}{2}\left\{(N+1)p+2\right\}}{Np+1}$$

$$\geq c(N+1) - \frac{vT + \frac{Nc}{2}\left\{(N+1)p+2\right\}}{Np+1}$$

$$= \frac{\frac{cp}{2}N(N+1) - vT + c}{Np+1}$$

$$\geq \frac{\frac{cp}{2}\frac{2Tv}{cp} - vT + c}{Np + 1}$$

$$= \frac{c}{Np+1}$$

$$\geq 0 .$$

Proof of (b). Fix $n < N$. If $n=o$, $X(o)$ is the only schedule. For $n > o$, suppose $\bar{x}$ is a schedule of n inspections such that $\max(g_{o\bar{x}}, \ldots, g_{n\bar{x}}) < \mu_{x(n)}$. Let

$$S_x = \sum_{k=o}^{n} g_{kx} .$$

Our supposition implies that $S_{\bar{x}} < S_{x(n)}$. From the last equation of (5) and those of (6), on adding, we have for all $x \in C_n$

$$p\,S_x + (1-p)g_{ox} = vT + \frac{nc}{2}(p(n+1) + 2).$$

However, since the right side of the above equation is constant, $S_{\bar{x}} < S_{x(n)}$ implies $g_{o\bar{x}} > g_{ox(n)}$, a contradiction. Therefore $x(n)$ is a minimax schedule in C_n. This, together with (a) and the finiteness of C^* proves (b).

Proof of (c). Notice that

$$\mu_{x(n)} - \mu_{x(n-1)} = \frac{cp^2n^2 + (2cp-cp^2)n + 2(c-pvT)}{2(np+1)\,((n-1)p+1)} , \quad n \leq N . \tag{7}$$

The sign of Eq. (7) is the sign of its numerator; therefore $\mu_{x(n)}$ is a nonincreasing function of n as long as the numerator of Eq. (7) is nonpositive. The numerator, however, is an increasing function of n and nonnegative for $n = N$. Therefore, the n such that $x^* = x(n)$ is given by Eq. (3) Q.E.D.

REFERENCE

[1] Barlow, Richard E., and Larry C. Hunter, "Mathematical Models for System Reliability," Engineering Report No. EDL-E35, Sylvania Electric Products Inc., Mountain View, California, 11 August 1959.

* * *

NOTES

PROBLEMS

It has been suggested that this journal might serve as a medium for publishing problems in the area of Logistics -- with the idea that other persons might be interested in those problems and might submit comments. Readers are invited to submit brief statements on applied and theoretical problems in Logistics. Address letters to Managing Editor, Naval Research Logistics Quarterly, Office of Naval Research, Washington 25, D. C.

ON PROFESSOR SCHELLING'S STRATEGY OF CONFLICT

John G. Cross
Princeton University

Professor Schelling's Strategy of Conflict [1] can be divided conceptually into three parts: an extensive verbal analysis of the nature of Threats, Commitments, and Promises and their uses in bargaining; a theory which may help one to understand the forces underlying the outcomes of bargaining situations; and a series of criticisms of Game Theory regarding its analysis of human behavior in conflict situations. In general, the terminology and utility matrices common in Game Theory are used, and in fact one may consider the Strategy of Conflict [1] to be an effort to bridge the gap between Game Theory analysis and actual developments under conflict.

The most fruitful part of the book is the excellent description of the use of threats, commitments, and promises, and the requirements which must be met before they can be put into operation successfully. Many provocative and entertaining examples are given, some of which shed considerable light on very ordinary experiences. The importance of commitments is emphasized, both as tactics in themselves (which can force opponents into taking the desired course of action), and as an important means of convincing an opponent that one intends to carry out a (perhaps mutually painful) threat if necessary.

Unfortunately, Schelling [1] fails even to consider an important objection to the use of such "strategies"[1] as threats and commitments: that of unfavorable psychological reactions to

[1] I hesitate to use the word "strategy" for fear of causing confusion: Schelling's definition [1] of a strategy as something concerned with the "exploitation of potential force" ([1], pp. 5, 160) is quite different from the game theoretical definition of a player's preplay plan which specifies what choices he will make in every possible situation which may arise during the course of a particular game (see [2] p. 79.)

them. The use of a threat, for instance, may very well antagonize an opponent (i.e., modify his utility values associated with the possible outcomes) so much that the threat does not work — to the disadvantage of both players, since the threat may have to be carried out. It would seem that such effects would have to be taken into account in any complete analysis of bargaining.

Schelling turns next to a discussion of some factors which can influence the specific outcomes of bargaining, particularly in cases of "tacit" bargaining in which the two players have no opportunity to communicate. He postulates that under such circumstances, the outcome is likely to be at some point of prominence (a "focal" point) of a mathematical or physical nature (such as a 50-50 split of a dollar, or a military division of land at some natural boundary). This analysis is primarily on an intuitive basis, in spite of an "informal" series of experiments which Schelling reports in order to substantiate his position. His analysis of why bargaining tends to settle at these focal points is rather sketchy — his conclusion that focal points provide obvious points at which a player can make a strong stand is accompanied by the remark that: "This proposition may seem intuitively plausible; it does to the author, and in any event, some kind of explanation is needed for the tendency to settle at focal points."[2]

In order to discuss Professor Schelling's criticisms of Game Theory, it is first necessary to distinguish between Game Theory proper and some of the more recent extensions to Bargaining Theory. The former, as originally presented by von Neumann and Morgenstern[3] is strictly an analysis of the range of possible outcomes of conflict situations, based on a few specific axioms. As it turns out, these axioms are usually not sufficiently restrictive to yield one unambiguous outcome for the game. For example, in the two-person, nonzero sum game, Game Theory concludes that "each player gets individually at least that amount which he can secure for himself, while the two get together precisely the maximum amount which they can secure together."[3] Thus, although we do know that the players will act in combination to extract the greatest possible gain (from "Nature"), we cannot say how this gain is divided between the two. The major contribution of Game Theory in this case is an elaborate framework into which new axioms can be introduced easily, to restrict further the number of outcomes. Bargaining theories consist of just such axioms, an example being Schelling's own assertion that focal points or prominences which are likely to have some subjective appeal to the players will affect, or even determine the outcome of the game. Unfortunately, most of these additional axioms involve much stronger assumptions about the behavior of the players, and therefore they are much more difficult to justify than those of Game Theory. Insofar as Schelling's criticism on page 119 [1] that Game Theory has been "pitched at too abstract a level of analysis" (referring to its inability to point to specific outcomes), we may reply that Game Theory never intended to give such results, its purpose being only to restrict the range of possible solutions as far as possible, using only its relatively weak assumptions, and leaving further restrictions to psychological analyses of bargaining.[4]

[2] See [1], p. 71.
[3] See [2], p. 555.
[4] See [2], p. 616.

A further issue stems from Schelling's assertions about Game Theory's "preoccupation with the zero sum game" and his many subsequent comparisons between nonzero sum games, bargaining, and the Game Theory analysis of two-person, zero sum games, particularly regarding the importance of communication between the two players: "There is probably no contrast more striking, in the comparison of the mixed motive [Schelling's term for a nonzero sum game] and the pure conflict (zero sum) game, then the significance of having one's own strategy found out and appreciated by the opponent. Hardly anything captures the spirit of the zero sum game quite so much as the importance of "not being found out" and of employing a mode of decision that is proof against deductive anticipation by the other player."[5]

It is difficult to understand how a theory which emphasizes the existence of incomplete and complete information, the "bluff,"[6] and the coalitions between players which must occur before the game can be "won," can be accused of basing its analysis on secrecy, or how a theory originally designed to treat economic problems "in which the sum of all payments, the total social product will in general not be zero and not even constant"[7] can be described as "preoccupied with the zero sum game." It is true that Game Theory bases its theoretical development on the two-person, zero sum game, and upon the mixed strategy, but not in the restricted sense in which Schelling has presented it. The "persons" in the Game Theory sense are merely groups of players who have combined, or who are classified together for the purposes of analysis (e.g., into the "winning" and the "losing" coalitions). These groups do use minimax strategies against one another when playing an n-person zero sum game. Furthermore, Game Theory is enabled to make use of its zero sum game analysis through the creation of a "fictitious player" who takes a purely passive role in the game, paying out whatever is demanded by the rest of the players according to the rules of the game — again only a mathematical device designed to simplify the analysis. In this manner, the two-person, nonzero sum game is treated as a three-person game in which the third player ("Nature") is exploited by the other two players in combination. The specific outcome, of course, depends (among other things) upon the rules of the game, communication, and bargaining between the two players. At no point in this analysis, however, do we have the two "real" players using minimax strategies against one another. Professor Schelling has quite inappropriately used the two-person, zero sum analysis as an example of the Game Theory approach to any two individuals in any game regardless of what other (perhaps "fictitious") players are also in it.

Schelling goes on to point out that the threats, promises, and commitments which he has demonstrated to be so important in bargaining are not explicitly considered in Game

---

[5] See [1], p. 160. At this point Schelling quotes the Theory (op. cit., p. 147) in a footnote, quoting a passage emphasizing the avoidance of being found out. This remark, however, refers only to certain zero sum games, and not to two players who have some interest in joint operations.

[6] Schelling's concept [1] of a "bluff," a misrepresentation or distortion of one's own value system (see [1] pp. 23, 33, 36), differs somewhat from the Game Theory definition of an impediment to the spread of information (concerning the state of affairs in a game in general or simply the strength of one's position) to an opponent. (See [2], pp. 54, 188.)

[7] See [2], p. 46.

Theory. This is because Game Theory generally treats the normalized form of the game, into which, as Schelling himself demonstrates on pages 150-156, these concepts can all be incorporated. Schelling complains that these tactics cannot be studied in this form, which is quite true, but Game Theory was not designed to study such things in themselves, but only their effects upon the possible outcomes of the game. Game Theory begins with the assumption that all existing alternatives including threats, promises, and the like, are already known, and are included in the normalized game.

Finally, Schelling criticizes Game Theory on the grounds that the normalization and randomization processes obliterate the "prominences" and "focal points" upon which his bargaining analysis depends.[8] For instance, he presents the matrix:

		Player B			
		I	II	III	IV
Player A	i	10, 10	0, 0	0, 0	0, 0
	ii	0, 0	10, 10	0, 0	0, 0
	iii	0, 0	0, 0	9, 9	0, 0
	iv	0, 0	0, 0	0, 0	10, 10

and states that if player A and player B are kept incommunicado, they will tend to choose iii III on the grounds that this payoff acts, through its uniqueness, as a focal point and thus provides a "clue to concert choices." Presumably, many other such "clues" which are inherent in the structure of the game have been lost through normalization. In the first place, it is important to note that the figures given in the Game Matrices are utilities and not physical units.[9] Therefore, in the many cases in which physical objects serve as the "prominences," these utilities cannot be used as focal points, and thus the prominences are kept outside the matrix. There is nothing to prevent us, however, from considering the fact that the strategies (I, II, III, IV, i, ii, iii, iv) possess prominences external to the matrix. In fact, Schelling's theory of focal points in which each player has some reason (which may or may not be purely intuitive) to assume that the other will choose a particular strategy, can be interpreted as the assignment of a probability distribution to each player's set of strategy choices by the other, who then acts accordingly. Hence the problem is reduced to one of decision making under risk – for each player. For instance, referring to the matrix above, player A may assume (subjectively) that player B will choose strategies I, II, III, IV in the ratio 1:1:2:1. The expected value of each of his strategies will be in the ratio:

$$\begin{aligned} &10 \text{ for strategy i} \\ &10 \text{ for strategy ii} \\ 2 \times 9 = &18 \text{ for strategy iii} \\ &10 \text{ for strategy iv} \end{aligned}$$

Hence he will chose iii. Whether player B assumes the same, or goes further and assumes that A makes the above calculation will make no difference – the outcome will still be iii III.

[8] See [1], pp. 96, 156.

[9] The Theory does in many cases assume money to be a direct measure of utility. (See [2], p. 79.)

This interpretation also demonstrates how a focal point may not always be the solution. Consider the matrix of payoffs:

		Player B			
		I	II	III	IV
Player A	i	3,3	0,0	0,0	0,0
	ii	0,0	3,3	0,0	0,0
	iii	0,0	0,0	1,1	0,0
	iv	0,0	0,0	0,0	3,3

If A again assumes that B will choose his strategies in the ratio 1:1:2:1, the expected values of his strategies will be in the ratio of:

3 for strategy i
3 for strategy ii
2 for strategy iii
3 for strategy iv

and A will not take iii, preferring a purely random strategy among i, ii, iv. All this simply serves to demonstrate how it is trivially simple to consider focal points within the context of Game Theory.

Schelling's assertion that Game Theory neglects the essential labels or prominences which make the above procedures possible evidently arises from his conclusion that Game and Bargaining Theories assume complete symmetry. It is true that certain forms of mathematical symmetry are assumed, but certainly such symmetry as would affect Schelling's results (i.e., that both players have exactly the same role in the game) is treated as a special case. Clearly, if we did assume that both players have identical strategic opportunities, or that both players have identical personalities and physical characteristics, many of Schelling's "labels" would be lost. It should be pointed out here, though, that these labels must be considered to be a part of the normalized game structure in the form of additional strategies, and that therefore the idealized "player" A or B does not correspond precisely to the player George Washington Roosevelt (who counts among his ancestors Capt. Miles Standish, Benjamin Franklin, and Henry VIII of England), who is running for State Senator against Joseph Glotz from East Howesville. Thus, only an assumption of identical strategic opportunities for the players A and B would preclude Schelling's analysis, and it is certain that in general, Game Theory makes no such assumption

ACKNOWLEDGMENT

The author is indebted to Dr. R. J. Aumann and Professor O. Morgenstern for their valuable advice and criticism of this article.

REFERENCES

[1] Thomas C. Schelling: The Strategy of Conflict. Harvard University Press, Cambridge, Mass., 1960.

[2] John von Neumann and Oskar Morgenstern: The Theory of Games and Economic Behavior. Princeton University Press, Princeton, N. J., 1953.

* * *

NEWS AND MEMORANDA

Readers are invited to submit to the Managing Editor items of general interest in the field of logistics

SYMPOSIUM ON OPTICAL CHARACTER RECOGNITION

A symposium on Optical Character Recognition will be held on 15, 16, and 17 January 1962. It will be jointly sponsored by the Information Systems Branch of the Office of Naval Research and the Research Information Center of the National Bureau of Standards. The Symposium will be held in the Department of the Interior Auditorium on C Street, between 18th and 19th Street, N.W., Washington, D. C.

The objective of the Symposium will be to bring together equipment specialists, scientists, and potential users who are interested in exploiting automatic character recognition techniques for a variety of applications. Major purposes are to explore, for mutual benefit, the present state of the art, operational requirements and the implications of current research. It is planned that the meeting will be composed of several sessions to cover, respectively, presentations on available equipment, directions in current research, representative requirements of potential users, and a forecast of future progress. The program for these sessions will consist of invited papers from individuals and organizations active in research and development efforts in the field, and from potential users. Where feasible, demonstrations of optical character recognition systems in operation in the Washington area will be arranged.

Attendance is open to all interested technical and management personnel. Further information and a preliminary Symposium program, when available, may be obtained by contacting:

Miss Josephine Leno
Code 430A
Office of Naval Research
Washington 25, D. C.
(Phone OXford 6-6213)

DONALD K. POLLOCK, ONR
BERNARD RADACK, BUSANDA
MARY ELIZABETH STEVENS, NBS
Symposium Co-Chairmen

SYMPOSIUM ON REDUNDANCY TECHNIQUES FOR COMPUTING SYSTEMS

A symposium on Redundancy Techniques for Computing Systems will be held on 6 and 7 February 1962; sponsored by the Information Systems Branch, Office of Naval Research. This Symposium will be held in the Department of the Interior Auditorium, C Street between 18th and 19th Street, N. W., Washington, D. C.

The objective of the Symposium is to focus attention toward new ideas, research and developments which may lead to the introduction of redundancy techniques into forthcoming computing systems. It is apparent that many information processing systems of the future will be of such size that conventional methods of fabrication and of maintenance will be completely infeasible. Computers for space applications have reliability requirements which are presently not achievable without use of redundancy. Thus, some form of logical redundancy must be introduced if practical systems are to be operated. Accordingly, it is planned to present in a single symposium a collection of related papers concerning suggested new techniques for the use of redundancy in computing systems. This should permit relative evaluation of such techniques with respect to their probable future utilization in various aspects of the computing field. The program will consist of papers invited from many of the organizations engaged in appropriate research and development activities.

Attendance is open to all interested technical personnel. Further information and a preliminary Symposium program, when available, may be obtained by contacting:

Miss Josephine Leno
Code 430A
Office of Naval Research
Washington 25, D. C.
(Phone OXford 6-6213)

RICHARD H. WILCOX
Symposium Co-Chairman

RECENT PUBLICATIONS

MODERN COMPUTING METHODS, Second Edition (completely revised) Philosophical Library, NEW YORK, pp. vi + 170, $6.

This is the second edition of the book initially written by L. Fox, E. T. Goodwin, J.G. L. Michel, F.W.J. Olver and J.H. Wilkinson. To write this edition, C.W. Clenshaw, D.W. Martin, and G. F. Miller have collaborated with the 2nd, 4th and 5th authors of the first edition.

The book contains 15 chapters together with a comprehensive bibliography. It is well written and contains illustrative examples. It is excellent reading for mathematicians and scientists who come in contact with computing.

The first five chapters discuss problems associated with different, computationally useful methods of solution of linear equations and latent root problems of matrices. Elimination methods e.g., Gaussian elimination and variants of it are discussed along with iterative processes, Jacobi's, Givens' and Householder's methods, for finding latent roots, with an emphasis on computation. Error analyses (e.g., for the Gaussian elimination method and for triangular decomposition) are given. In chapter six, methods of finding zeros of polynomials are discussed (e.g., Newton's rule, Bairstow's process and other iterative methods). The remaining chapters are on finite differences, Chebyshev series ordinary and partial differential equations, use of recurrence relations for evaluating limits, the evaluation of integrals and the tabulation of mathematical functions.

The book treats the topics covered in an interesting fashion with greatest concern for computational aspects but often developing the required theory for background material of which of particular interest are the presentations on matrices and on hyperbolic partial differential equations.

The subject of constrained maxima and minima is much in demand and occupies the attention of many people. It is not discussed in the book nor does the book contain general philosophical background which points out the nature of classes of general problems and their associated computational subtleties. The readers of the Quarterly will find it a welcome addition to a subject much in demand, particularly for its broad coverage of topics and treatment of computational problems in an appealing way.

(Reviewed by Thomas L. Saaty)

ELEMENTS OF MATHEMATICAL STATISTICS, by Howard W. Alexander. John Wiley & Sons, Inc., New York, 1961. xi + 367 pp.

In spite of the publication of many books on statistics, both students and teachers of statistics are having difficulty in finding a textbook which is appropriate to the mathematical

backgrounds of students. In this sense we appreciate the author's intention to "make mathematical statistics really accessible to the college senior or junior whose background includes a typical calculus course."

In this book, Professor Alexander takes up adequately almost all statistical methods—but he does not refer to the concept of decision theory—without showing any of the bias which is apparent in several books, according to the authors' predilections. I think this impartiality, as shown in the following chapters in Alexander's book, is an important qualification for a good textbook.

Chapter 1 Discrete Distributions
2 Continuous Distributions
3 Sampling Distributions I
4 Statistical Inference
5 Sampling Distributions II
6 Regression and Correlation
7 Analysis of Variance
Appendix Matrices

His introduction of the concept of probability is especially excellent. In teaching statistics, we have difficulty at times in proceeding from the discrete distribution to the continuous distribution. The author of this book gives a very natural and understandable explanation of this point.

Generally speaking, I think the author has succeeded in realizing his purpose. Nonetheless, I have some questions as to his presentation from the pedagogical point of view:

1. When we consider the importance of the concept of the expected value in many applications of statistical methods, it seems to me necessary to give an empirical explanation of this concept, in relation to the concept of probability as it is given in this book. We find many people who are applying this concept superficially to situations where repetition of the trial is not possible.

2. In testing hypotheses, the author does not distinguish clearly between simple hypotheses and composite hypotheses. If we were to follow the traditional approach to statistics—as the author has, as shown in his neglect of the explanation of the statistical decision function—it seems necessary to make clear the difference between these two kinds of hypotheses.

3. I am afraid that Alexander's explanation of the difference between the concepts of the confidence coefficient and probability will confuse students.

4. It seems necessary to give an empirical explanation of the meaning of the unbiasedness in the point estimation. This point is related to my question 1.

5. When we consider the powerful applicability of the least squares method, it seems unreasonable to omit the proof of the unbiasedness of the least squares estimates without assuming normality, when the proof of the distributions of these estimates under the assumption of normality is given.

Notwithstanding these questions, I think this is a good textbook on statistics.

(Reviewed by Koichi Miyasawa)

INDEX OF VOLUME 8

	pages
ARROW, K. J., Hurwicz, Leonid, and Uzawa, Hirofumi, Constraint Qualifications in Maximization Problems	175-191
BALINSKI, Michel L., Fixed-Cost Transportation Problems	41- 54
BARANKIN, E. W., A Delivery-Lag Inventory Model With an Emergency Provision (The Single-Period Case)	285-311
BOVAIRD, R. L., and Zagor, H. I., Lognormal Distribution and Maintainability in Support Systems Research	343-356
CALDWELL, W. V., Coombs, C. H., Schoeffler, M.S., and Thrall, R. M., A Model for Evaluating the Output of Intelligence Systems	25- 40
CHARNES, A., Cooper, W. W., and Miller, M., Dyadic Programs and Subdual Methods	1- 23
COOMBS, C. H., Caldwell, W. V., Schoeffler, M. S., and Thrall, R. M. A Model for Evaluating the Output of Intelligence Systems	25- 40
COOPER, W. W., Charnes, A., and Miller, M., Dyadic Programs and Subdual Methods	1- 23
CROSS, John G., On Professor Schelling's Strategy of Conflict	421-425
DERMAN, C., On Minimax Surveillance Schedules	415-419
ECCLES, H. E., Logistics: Conditio sine qua non for NATO Defense	111-116
FLOOD, Merrill M., A Transportation Algorithm and Code	257-276
FUKUDA, Yoichiro, Optimal Disposal Policies	221-227
GARG, R. C., and Mohan, C., Decision on Retention of Excess Stock Following a Normal Probability Law of Obsolescence and Deterioration	229-234
GLUSS, Brian, Approximately Optimal One-Dimensional Search Policies in Which Search Costs Vary Through Time	277-283
GLUSS, Brian, An Alternative Solution to the "Lost at Sea" Problem	117-121
GLUSS, Brian, The Minimax Path in a Search for a Circle in a Plane	357-360
HADLEY, G., and Whitin, T. M., Budget Constraints in Logistics Models	215-220
HADLEY, G., and Whitin, T. M., A Model for Procurement, Allocation, and Redistribution for Low Demand Items	395-414
HUNTER, Larry and Proschan, Frank, Replacement When Constant Failure Rate Precedes Wearout	127-136
HURWICZ, Leonid, Arrow, K. J., and Uzawa, Hirofumi, Constraint Qualifications in Maximization Problems	175-191

pages

LOMBARDI, Lionello, A Direct Method for the Computation of Waiting Times 193-197

MANNE, A. S., On the Timing of Development Expenditures and the Retirement of Military Equipment 235-243

MEYER, R. F., and Wolfe, H. B., The Organization and Operation of a Taxi Fleet 137-150

MILLER, M., Charnes, A., and Cooper, W. W., Dyadic Programs and Subdual Methods 1- 23

MOHAN, C., and Garg, R. C., Decision on Retention of Excess Stock Following a Normal Probability Law of Obsolescence and Deterioration 229-234

PROSCHAN, Frank, and Hunter, Larry, Replacement When Constant Failure Rate Precedes Wearout 127-136

PRUITT, W. E., A Class of Dynamic Games 55- 78

ROSENBLATT, Judah, Statistical Aspects of Collision Warning System Design Under the Assumption of Constant Velocity Courses 317-341

SALVESON, M. E., Principles of Dynamic Weapon Systems Programming 79-110

SCHOEFFLER, M. S., Caldwell, W. V., Coombs, C. H., and Thrall, R. M., A Model for Evaluating the Output of Intelligence Systems 25- 40

SHAPLEY, L. S., On Network Flow Functions 151-158

SHARPE, W. F., Aircraft Compartment Design Criteria for the Army Deployment Mission 381-394

SHOEMAKER, R. M., Principles of Logistics – A Provisional Definition. 159-173

THRALL, R. M., Caldwell, W. V., Coombs, C. H., and Schoeffler, M. S., A Model for Evaluating the Output of Intelligence Systems 25- 40

UZAWA, Hirofumi, Arrow, K. J., and Hurwicz, Leonid, Constraint Qualifications in Maximization Problems 175-191

WEISS, Lionel, "Hedging" on Statistical Assumptions 207-213

WEISS, Lionel, On the Estimation of Scale Parameters 245-256

WHELAN, D. W., SC, USN, Material Logistic Support of Weapon Systems 361-375

WHITIN, T. M., and Hadley, G., Budget Constraints in Logistics Models 215-220

WHITIN, T. M., and Hadley, G., A Model for Procurement, Allocation, and Redistribution for Low Demand Items 395-414

WOLFE, H. B., and Meyer, R. F., The Organization and Operation of a Taxi Fleet 137-150

WURTELE, Zivia S., A Note on Pyramid Building 377-379

ZAGOR, H. I., and Bovaird, R. L., Lognormal Distribution and Maintainability in Support Systems Research 343-356

☆ U.S. GOVERNMENT PRINTING OFFICE : 1962 O - 633478

INFORMATION FOR CONTRIBUTORS

The NAVAL RESEARCH LOGISTICS QUARTERLY is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Manuscripts and other items for publication should be sent to The Managing Editor, NAVAL RESEARCH LOGISTICS QUARTERLY, Office of Naval Research, Washington 25, D. C. Each manuscript which is considered to be suitable material for the QUARTERLY is sent to one or more referees.

Manuscripts submitted for publication should be typewritten, double-spaced, and the author should retain a copy. Refereeing may be expedited if an extra copy of the manuscript is submitted with the original.

A short abstract (not over 400 words) should accompany each manuscript. This will appear at the head of the published paper in the QUARTERLY.

There is no authorization for compensation to authors for papers which have been accepted for publication. Authors will receive 50 reprints of their published papers.

Readers are invited to submit to the Managing Editor items of general interest in the field of logistics, for possible publication in the NEWS AND MEMORANDA or NOTES sections of the QUARTERLY.